



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification⁵ : C12N 15/11, C12Q 1/68	A2	(11) International Publication Number: WO 93/16178 (43) International Publication Date: 19 August 1993 (19.08.93)
(21) International Application Number: PCT/US93/01294 (22) International Filing Date: 12 February 1993 (12.02.93) (30) Priority data: 07/837,195 12 February 1992 (12.02.92) US (71) Applicant: THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Washington, DC (US). (72) Inventors: VENTER, Craig, J. ; 1718 Nordic Hill Circle, Silver Spring, MD 20906 (US). ADAMS, Mark, D. ; 12812 Sage Terrace, Germantown, MD 20874 (US). MORENO, Ruben, F. ; 14415 Coral Gables Way, North Potomac, MD 20878 (US).		(74) Agents: ALTMAN, Daniel, E. et al.; Knobbe, Martens, Olson and Bear, 620 Newport Center Drive, 16th Floor, Newport Beach, CA 92660 (US). (81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: SEQUENCES CHARACTERISTIC OF HUMAN GENE TRANSCRIPTION PRODUCT (57) Abstract Partial and complete human cDNA and genomic sequences corresponding to particular expressed sequence tags (ESTs). The ESTs are cDNA sequences that are generally between 150 and 500 base pairs in length, are derived from human brain cDNA libraries, correspond to genes transcribed in human brain, and have base sequences identified herein as SEQ ID NOS: 1-2421.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SU	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

**SEQUENCES CHARACTERISTIC OF HUMAN GENE TRANSCRIPTION
PRODUCT**

5

Technical Field

The present invention relates to newly identified polynucleotide sequences corresponding to transcription products of human genes, and to complete gene sequences associated therewith.

10

Background

This invention relates to human genes. Identification and sequencing of human genes is a major goal of modern scientific research. The sequence of human genes is more than just a scientific curiosity. For example, by identifying genes and determining their sequences, scientists have been able to make large quantities of valuable human "gene products." These include human insulin, interferon, Factor VIII, tumor necrosis factor, human growth hormone, tissue plasminogen activator, and numerous other compounds. Additionally, knowledge of gene sequences can provide the key to treatment or cure of genetic diseases (such as muscular dystrophy and cystic fibrosis). The present invention represents a quantum leap forward in mankind's knowledge of human gene sequences.

20

25

There are several basic concepts of molecular biology which figure prominently in the invention. A brief explanation of those concepts follows. Additional background information and definitions for scientific terms can be found

30

-2-

in the literature. See, for example, "Glossary of Genetics, Classical and Molecular" by R. Rieger, A. Michaelis, and M.M. Green (Fifth Edition, Springer-Verlag, New York (1991)). The contents of this and other publications cited in the specification are incorporated by reference herein.

At an initial level, the present invention is based on identification and characterization of gene segments. Genes are the basic units of inheritance. Each gene is a string of connected bases called nucleotides. Most genes are formed of deoxyribonucleic acid, DNA. (Some viruses contain genes of ribonucleic acid, RNA.) The genetic information resides in the particular sequence in which the bases are arranged. A short sequence of nucleotides is often called a polynucleotide or an oligonucleotide.

Like genes, polypeptides are built from long strings of individual units. These units are amino acids. The nucleotide sequence of a gene tells the cell the sequence in which to arrange the amino acids to make the polypeptide encoded by that gene. In general, chains of up to about 200 amino acids are called polypeptides, while proteins are larger molecules made up of polypeptide subunits; both types of molecules are referred to generally herein as polypeptides. A triplet of nucleotides (codon) in DNA codes for each amino acid or signals the beginning or end of the message (anticodon). The term codon is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the original DNA sequence is transcribed.

Generally, enzymes in the cell transcribe the permanent DNA of the gene into a temporary RNA copy, called messenger RNA or mRNA. The mRNA, in turn, can be translated into a polypeptide by the cell. This entire process is called gene expression, and the polypeptide is the gene product encoded by the gene.

Scientists have previously discovered how to reverse the transcription process and copy mRNA back into DNA using an

-3-

enzyme called reverse transcriptase. The resulting is called complementary DNA, or cDNA. This is schematically shown in the single Figure. When substantially all of the mRNA from one cell or tissue is converted to cDNA at once and cloned into multiple copies of a recombinant vector to allow replication and manipulation in the laboratory, the result is called a cDNA library.

The various types of genes include those which code for polypeptides, those which are transcribed into RNA but are not translated into polypeptides, and those whose functional significance does not demand that they be transcribed at all. Most genes are found on large molecules of DNA located in chromosomes. Double stranded cDNA carries all the information of a gene. Each base of the first strand is joined to a complementary base (hybridized) in the second strand. The linear DNA molecules in chromosomes have thousands of genes distributed along their length. Chromosomes include both coding regions (coding for polypeptides) and noncoding regions; the coding regions represent only about three percent of the total chromosome sequence.

An individual gene has regulatory regions that include a promoter which directs expression of the gene, a coding region which can code for a polypeptide, and a termination signal. The regulatory DNA sequence is usually a noncoding region that determines if, where, when, and at what level a particular gene is expressed.

The coding regions of many genes are discontinuous, with coding sequences (exons) alternating with noncoding regions (introns). The final mRNA copy of the gene does not include these introns (which can be much longer than the coding region itself), although it does contain certain untranslated regions that usually do not code for the polynucleotide gene product. Untranslated sequences at the beginning and end of the mRNA are known as 5'- and 3'-untranslated regions,

respectively. This nomenclature reflects the orientation of the nucleotide constituents of the mRNA.

5 A cDNA is a DNA copy of a messenger RNA, which contains all of the exons of a gene. The cDNA can be thought of as having three parts: an untranslated 5' leader, an uninterrupted polypeptide-coding sequence, and a 3' untranslated region. The untranslated leader and trailing sequences are important for initiation of translation, mRNA stability, and other functions. The untranslated leader and
10 trailing sequences are called 5'- and 3'-untranslated sequences, respectively. The 3' untranslated sequence is usually longer than the 5' untranslated leader, and can be longer than the polypeptide-coding sequence. The untranslated regions typically have many, randomly-
15 distributed stop codons, and do not display the nonrandom base arrangements found in coding sequences. The 5'-untranslated sequence is relatively short, generally between 20 and 200 bases. The 3'-untranslated sequence is often many times longer, up to several thousand bases.

20 The translated or coding sequence begins with a translational start codon (AUG or GUG) and ends with a translational stop codon (UAA, UGA, or UAG). Generally, translation begins at the first "start" codon on the mRNA and proceeds to the first "stop" codon. Coding sequences can be
25 distinguished by their nonrandom distribution of bases; numerous computer algorithms have been developed to distinguish coding from noncoding regions in this way.

Human DNA differs from person to person. No two persons (except perhaps identical twins) have identical DNA. While
30 the differences, called allelic variations or polymorphisms, are slight on a molecular level, they account for most of the physical and other observable differences between individuals. It has been estimated that approximately 14 million sequence polymorphism differences exist between
35 individuals.

-5-

The ability of one strand of DNA to attach or hybridize to a complementary strand has already been exploited for several purposes. For example, small pieces of DNA (15 to 25 base pairs long) can be made which will hybridize to longer strands of DNA which have a complementary sequence. These short "primers" can be selected such that they hybridize to a specific, unique location on the longer strand. Once the primers have hybridized to their target on the DNA, the polymerase chain reaction (PCR) can be employed to generate millions of copies of (or amplify) the particular segment of DNA between the locations to which two primers are bound. Briefly, this technique allows amplification of a DNA region situated between two convergent primers, using oligonucleotide primers that hybridize to opposite strands. Primer extension proceeds inward across the region between the two primers, and the product of DNA synthesis of one primer serves as a template for the other primer. Repeated cycles of DNA denaturation, annealing of primers, and extension result in an exponential increase in the number of copies of the region bounded by the primers.

Similarly, a labeled segment of single-stranded DNA can be hybridized to a longer DNA sequence, such as a chromosome, to mark a specific location on the longer sequence. Segments of DNA 50 bases long or longer that hybridize to a unique DNA location in the human genome are extremely unlikely to hybridize elsewhere in the human genome.

The Human Genome Project is an effort to sequence all human DNA (the human genome). The human genome is estimated to comprise 50,000 - 100,000 genes, up to 30,000 of which might be expressed in the brain (Sutcliffe, *Ann. Rev. Neurosci.* 11:157 (1988)). Once dedicated human chromosome sequencing begins in three to five years, it was expected that 12-15 years will be required to complete the sequence of the genome (Report of the Ad Hoc Program Advisory Committee on Complex Genomes, Reston, Va., Feb. 1988, D. Baltimore Ed. (NIH, Bethesda, Md, 1988)). At that rate, the majority of

-6-

human genes would remain unknown for at least the next decade. The present invention can greatly accelerate the pace at which human genes can be identified and mapped. Most gene researchers, in conjunction with publication of their results in this field, submit sequence data to the GenBank database. Prior to the present invention, GenBank listed the sequences of only a few thousand human genes and less than two hundred human brain mRNAs (GenBank Release 66.0, December, 1990).

The role of sequencing complementary DNA (cDNA), reverse transcribed from mRNA, as a part of the human genome project has been vigorously debated since the idea of determining the complete nucleotide sequence of humans first surfaced. The coding sequence of all human genes represents most of the information content of the genome, but only 3-5% of the total DNA. In contrast, cDNA (which is only made from the transcription product of active genes) is one-half to three-fourths (the remainder being 5'- and 3'-untranslated sequence) meaningful genetic information. Thus, some have argued that cDNA sequencing should take precedence over genomic sequencing (Brenner, CIBA Found. Symp. 149:6 (1990)). However, until now, such arguments have not been heeded.

Genomic sequencing proponents have argued the difficulty of finding every mRNA expressed in all tissues, cell types, and developmental states, and that much valuable information from intronic and intergenic regions, including control and regulatory sequences, will be missed by cDNA sequencing. (Report of the Committee on Mapping and Sequencing the Human Genome, National Research Council (National Academy Press, Washington, D.C. 1988)). Further, sequencing of transcribed regions of the genome using cDNA libraries has heretofore been considered impractical or unsatisfactory. Libraries of cDNA were believed to be dominated by repetitive elements, mitochondrial genes, ribosomal RNA genes, and other nuclear genes comprising common or housekeeping sequences. It was believed that cDNA libraries would provide few sequences

-7-

corresponding to structural and regulatory polypeptides or peptides. See, for example, Putney, et al., *Nature* 302:718-721 (1983). Putney, et al. sequenced over 150 clones from a rabbit muscle cDNA library and identified clones for 13 of the 19 known muscle polypeptides, including one new isotype but no unknown coding sequences.

Another perceived drawback of cDNA sequencing was that some mRNAs are abundant, and some are rare. The cellular quantities of mRNA from various genes can vary by several orders of magnitude. This led critics to believe that most information obtained from cDNA sequencing would be repetitious and useless.

The present invention demonstrates that, despite such skepticism, cDNA sequencing now provides a rapid method for obtaining enormous amounts of valuable genetic information and DNA products of great utility for the biotechnology and pharmaceutical industries. Not only can many distinct cDNAs be isolated and sequenced, even partial cDNAs can be used, with conventional, well-understood methods, to isolate entire genes, and to determine the chromosomal locations and biological functions of these genes. As is demonstrated here, fragments of only a few hundred bases are sufficient, in many cases, to identify the probable function of a new human gene if it is similar in structure to a gene from another animal, or from plants or bacteria. Similarly, even fragments of untranslated regions of a cDNA can be used to: i) isolate the coding sequence of the cDNA; ii) isolate the complete gene; iii) determine the position of the gene on a human chromosome, and hence the potential of the gene to cause a human genetic disease; and iv) determine the function of the gene by means of experiments in which the function of the native gene is disrupted by the addition of a short DNA fragment to the cell, e.g., using triple helix or antisense probes.

Because coding regions comprise such a small portion of the human genome, identification and mapping of transcribed

-8-

regions and coding regions of chromosomes is of significant interest. There is a corresponding need for reagents for identifying and marking coding regions and transcribed regions of chromosomes. Furthermore, such human sequences are valuable for chromosome mapping, human identification, identification of tissue type and origin, forensic identification, and locating disease-associated genes (i.e., genes that are associated with an inherited human disease, whether through mutation, deletion, or faulty gene expression) on the chromosome.

SUMMARY OF THE INVENTION

Contrary to the expectations of the scientific community, cDNA screening and sequencing techniques have now been used to discover a large number of heretofore unknown human genes. Disclosed herein are over 2,400 new human polynucleotide sequences. These sequences could represent up to 5% of all human genes. The novelty of these sequences has been established through comparison to both nucleotide sequence databases and amino acid sequence databases. Surprisingly, over 80% of the sequences generated were unrelated to any sequences previously described in the literature.

The sequences of the present invention were ascertained using a fast approach to cDNA characterization. This approach could facilitate the tagging of most expressed human genes within a few years at a fraction of the cost of complete genomic sequencing, provide new genetic markers, provide new DNA-based therapeutics and diagnostics, and provide other valuable nucleotide reagents.

The sequences disclosed herein, styled Expressed Sequence Tags ("ESTs"), are markers for human genes actually transcribed *in vivo*. Techniques are disclosed for using these ESTs to obtain the full coding region of the corresponding gene. The use of ESTs, complete coding sequences, or fragments thereof for marking chromosomes, for

-9-

mapping locations of expressed genes on chromosomes, for individual or forensic identification, for mapping locations of disease-associated genes, for identification of tissue type, and for preparation of antisense sequences, probes, and constructs is discussed in detail below. Unlike the random genomic DNA sequence tagged sites (STSs) (Olson et al., *Science* 245:1434 (1989)), ESTs point directly to expressed genes.

Various aspects of the present invention thus include the individual ESTs, corresponding partial and complete cDNA, genomic DNA, mRNA, antisense strands, triple helix probes, PCR primers, coding regions, and constructs. Also, where one skilled in the art is enabled by this specification to prepare expression vectors and polypeptide expression products, they are also within the scope of the present invention, along with antibodies, especially monoclonal antibodies, to such expression products.

BRIEF DESCRIPTION OF THE DRAWING

The single drawing Figure schematically illustrates the progression from chromosome to gene to mRNA to cDNA.

DETAILED DESCRIPTION OF THE INVENTION

The detailed description that follows provides not only the actual sequence of each new EST, but also explains how the ESTs were obtained, how to obtain the corresponding complete cDNA sequence and the corresponding genomic DNA sequence, how to make DNA constructs from the ESTs and corresponding sequences, how to use those sequences as reagents in molecular biology and other fields, how to produce gene products from the ESTs and corresponding sequences and antibodies to those gene products, and the functional categories of many ESTs and corresponding genes. Furthermore, numerous actual working examples and predictive

-10-

examples are provided to demonstrate and exemplify numerous aspects of the invention.

I. ESTs from cDNA Libraries

5 The sequences of the present invention were isolated from commercially available and custom made cDNA libraries using a rapid screening and sequencing technique. In general, the method comprises applying conventional automated DNA sequencing technology to screening clones, advantageously
10 randomly selected clones, from a cDNA library. Preferably, the library is initially "enriched" through removal of ribosomal sequences and other common sequences prior to clone selection. According to the present method, ESTs are generated from partial DNA sequencing of the selected clones.
15 The ESTs of the present invention were generated using low redundancy of sequencing, typically a single sequencing reaction. While single sequencing reactions may have an accuracy as low as 97%, this nevertheless provides sufficient fidelity for identification of the sequence and design of PCR
20 primers.

 Most human genes can be identified by EST sequencing from libraries of cDNA copies of messenger RNAs. However, some genes are expressed only at specific times during embryonic development, or only in small amounts in a few
25 specific cell types. Other genes have mRNAs that are degraded very quickly by the cell in which they are expressed. If any of these are the case, transcripts of the gene will not be represented in cDNA libraries so the gene will not be identifiable by EST sequencing. A new method
30 called "exon amplification", however, can be used to isolate and identify transcripts of such genes.

 Exon amplification works by artificially expressing part or all of a gene that is contained in a cloned fragment of genomic DNA such as a cosmid or yeast artificial chromosome
35 (YAC). The gene is cloned into a special vector, designed at MIT, that uses control elements from virus genes to express

-11-

the protein-coding exons of the human gene of interest. Exon trapping shows considerable promise as a general technique for identifying those genes in the human genome that cannot be found by cDNA cloning and EST sequencing. Exon amplification will also be useful for identifying the genes in regions of genomic DNA to which disease genes have been mapped. The exon amplification method can be used directly with the cosmid and YAC clones from human chromosomes that are being obtained by both NIH and DOE supported human genome centers. ESTs comprise DNA sequences corresponding to a portion of nuclear encoded messenger RNA. An EST is of sufficient length to permit: (1) amplification of the specific sequence from a cDNA library, e.g., by polymerase chain reaction (PCR); (2) use of a synthetic polynucleotide corresponding to a partial or complete sequence of the EST as a hybridization probe of a cDNA library, generally having 30 - 50 base pairs; or (3) unique designation of the pure cDNA clone from which the EST was derived (the EST clone) for use as a hybridization probe of a cDNA library. Preferably, EST-derived primer pairs and sequences amplify or detectably hybridize to a sequence from a genomic library.

It has been found that sufficient information is contained in the 150-400 base ESTs from one sequencing run to effect preliminary identification and exact chromosome mapping. Accordingly, the ESTs disclosed herein are generally at least 150 base pairs in length. The length of an EST is determined by the quality of sequencing data and the length of the cloned cDNA. Raw data from the automated sequencers is edited to remove low quality sequence at the end of the sequencing run. High quality sequences (usually a result of sequencing templates without excessive salt contamination) generally give about 400 bp of reliable sequence data; other sequences give fewer bases of reliable data. A 150 bp EST is long enough to be translated into a 50 amino acid peptide sequence. This length is sufficient to observe similarities when they exist in a database search. Furthermore, 150 bp is

-12-

long enough to design PCR primers from each end of the sequence to amplify the complete EST. Sequences shorter than 150 bp are difficult to purify and use following PCR amplification. Furthermore, a 150 bp polynucleotide is likely to give a very strong signal with low background in a screen of a genomic library.

Finally, it is highly unlikely that a sequence of the same 150 bp exists in any genes in the genome besides the one tagged by the EST. Some closely related gene family members have very similar nucleotide sequences, but no examples of pairs of human genes with long segments of identical sequence have been reported to date. For instance, there are three known β -tubulin genes in humans. Several ESTs were found that matched one or another of these tubulin genes, but several new members of this gene family were also found and could be clearly distinguished from the three known members. ESTs that match perfectly to several different genes can be detected by hybridizing to chromosomes: if many chromosomal loci are observed, the sequence (or a close variant) is present in more than one gene. This problem can be circumvented by using the 3'-untranslated part of the cDNA alone as a probe for the chromosomal location or for the full-length cDNA or gene. The 3'-untranslated region is more likely to be unique within gene families, since there is no evolutionary pressure to conserve a coding function of this region of the mRNA.

As demonstrated in the Examples that follow, ESTs can be used to map the expressed sequence to a particular chromosome. In addition, ESTs can be expanded to provide the full coding regions, as detailed below. In this manner, previously unknown genes can be identified.

While a variety of cDNA libraries can be used to obtain ESTs, human brain cDNA libraries are exemplified and represent a preferred embodiment. Suitable cDNA libraries can be freshly prepared or obtained commercially, e.g., as shown in Examples 1, 2, and 11. The cDNA libraries from the

-13-

desired tissue are preferably preprocessed by conventional techniques to reduce repeated sequencing of high and intermediate abundance clones and to maximize the chances of finding rare messages from specific cell populations. Preferably, preprocessing includes the use of defined composition prescreening probes, e.g., cDNA corresponding to mitochondria, abundant sequences, ribosomes, actins, myelin basic polypeptides, or any other known high abundance peptide; these prescreening probes used for preprocessing are generally derived from known ESTs. Other useful preprocessing techniques include subtraction, which preferentially reduces the population of certain sequences in the library (e.g., see A. Swaroop et al., *Nucl. Acids Res.* 19, 1954 (1991)), and normalization, which results in all sequences being represented in approximately equal proportions in the library (Patanjali et al, *Proc. Natl. Acad. Sci. USA* 88:1943 (1991)).

The cDNA libraries used in the present method will ideally use directional cloning methods so that either the 5' end of the cDNA (likely to contain coding sequence) or the 3' end (likely to be a non-coding sequence) can be selectively obtained."

Libraries of cDNA can also be generated from recombinant expression of genomic DNA. After they are amplified, ESTs can be obtained and sequenced, e.g., as illustrated in Example 11.

The sequences of the present invention include the specific sequences set forth in the Sequence Listing and designated SEQ ID NO: 1 - SEQ ID NO: 2412. In one aspect of this embodiment, the invention relates to those sequences of SEQ ID NOS: 1 - 2412 that comprise the cDNA coding sequences for polypeptides having less than 95% identity with known amino acid sequences (see Table 2) and more preferably less than 90% or 85% identity. In a second aspect, the invention relates to those sequences of SEQ ID NOS: 1 - 2412 that encode polypeptides having no similarity to known amino acid

-14-

sequences (see Examples that follow). Precisely because they do not contain coding regions and are therefore more unique in their sequence structures, those sequences which meet neither of the preceding criteria can be most useful and are generally preferred for mapping.

Consistent with the NIH mission and its responsibilities to disseminate knowledge and share the tangible fruits of its research, the present inventors have taken a number of steps to facilitate sequence data and clone availability. All EST sequences have been submitted to GenBank (representing an addition equivalent to 7% of the human nucleotides in Release 69 of GenBank, September 1991). The corresponding cDNA clones have been submitted to the American Type Culture Collection and information on clones and sequences has been submitted to the Genome Data Base (Pearson, P. Nucl. Acids Res. 19 (Suppl.): 2237-9 (1991)).

II. Complete Coding Sequences from ESTs

The ESTs of the present invention generally represent relatively small coding regions or untranslated regions of human genes. Although most of these sequences do not code for a complete gene product, the ESTs of the present invention are highly specific markers for the corresponding complete coding regions. The ESTs are of sufficient length that they will hybridize, under stringent conditions, only with DNA for that gene to which they correspond. Suitably stringent conditions comprise conditions, for example, where at least 95%, preferably at least 97% or 98% identity (base pairing), is required for hybridization. This property permits use of the EST to isolate the entire coding region and even the entire sequence. Therefore, only routine laboratory work is necessary to parlay the unique EST sequence into the corresponding unique complete gene sequence.

Thus, each of the ESTs of the present invention "corresponds" to a particular unique human gene. Knowledge

-15-

of the EST sequence permits routine isolation and sequencing of the complete coding sequence of the corresponding gene. The complete coding sequence is present in a full-length cDNA clone as well as in the gene carried on genomic clones. Therefore, each EST "corresponds" to a cDNA (from which the EST was derived), a complete genomic gene sequence, a polypeptide coding region (which can be obtained either from the cDNA or genomic DNA), and a polypeptide or amino acid sequence encoded by that region.

The first step in determining where an EST is located in the cDNA is to analyze the EST for the presence of coding sequence, e.g., as described in Example 14. The CRM program predicts the extent and orientation of the coding region of a sequence. Based on this information, one can infer the presence of start or stop codons within a sequence and whether the sequence is completely coding or completely non-coding. If start or stop codons are present, then the EST can cover both part of the 5'-untranslated or 3'-untranslated part of the mRNA (respectively) as well as part of the coding sequence. If no coding sequence is present, it is likely that the EST is derived from the 3'-untranslated sequence due to its longer length and the fact that most cDNA library construction methods are biased toward the 3' end of the mRNA.

One general procedure for obtaining complete sequences from ESTs is as follows:

1. Purify selected human DNA from an EST clone (the cDNA clone that was sequenced to give the EST), e.g., by endonuclease digestion using ECOR1, gel electrophoresis, and isolation of the aforementioned clone by removal from low-melting agarose gel.

2. Radiolabel the isolated insert DNA, e.g., with ³²P labels, preferably by nick translation or random primer labeling.

3. Use the labeled EST insert as a probe to screen a lambda phage cDNA library or a plasmid cDNA library.

-16-

4. Identify colonies containing clones related to the probe cDNA and purify them by known purification methods.

5. Nucleotide sequence the ends of the newly purified clones to identify full length sequences.

5 6. Perform complete sequencing of full length clones by Exonuclease III digestion or primer walking. Northern blots of the mRNA from various tissues using at least part of the EST clone as a probe can optionally be performed to check the size of the mRNA against that of the purported full
10 length cDNA.

 An EST is a specific tag for a messenger RNA molecule. The complete sequence of that messenger RNA, in the form of cDNA, can be determined using the EST as a probe to identify a cDNA clone corresponding to a full-length transcript,
15 followed by sequencing of that clone. The EST or the full-length cDNA clone can also be used as a probe to identify a genomic clone or clones that contain the complete gene including regulatory and promoter regions, exons, and introns.

20 ESTs are used as probes to identify the cDNA clones from which an EST was derived. ESTs, or portions thereof, can be nick-translated or end-labelled with P^{32} using polynucleotide kinase using labelling methods known to those with skill in the art. (*Basic Methods in Molecular Biology*, L.G. Davis, M.D.
25 Dibner, and J.F. Battey, ed., Elsevier Press, NY, 1986). The lambda library can be directly screened with the labelled ESTs of interest or the library can be converted en masse to pBluescript (Stratagene, La Jolla, California) to facilitate bacterial colony screening. Both methods are well known in
30 the art. Briefly, filters with bacterial colonies containing the library in pBluescript or bacterial lawns containing lambda plaques are denatured and the DNA is fixed to the filters. The filters are hybridized with the labelled probe using hybridization conditions described by Davis et al. The
35 ESTs, cloned into lambda or pBluescript, can be used as positive controls to assess background binding and to adjust

-17-

the hybridization and washing stringencies necessary for accurate clone identification. The resulting autoradiograms are compared to duplicate plates of colonies or plaques; each exposed spot corresponds to a positive colony or plaque. The colonies or plaques are selected, expanded and the DNA is isolated from the colonies for further analysis and sequencing.

The ESTs can additionally be used to screen Northern blots of mRNA obtained from various tissues or cell cultures, including the tissue of origin of the EST clone. Northern analysis will most often produce one to several positive bands. The bands can be selected for further study based on the predicted size of the mRNA.

Positive cDNA clones in phage lambda are analyzed to determine the amount of additional sequence they contain using PCR with one primer from the EST and the other primer from the vector. Clones with a larger vector-insert PCR product than the original EST clone are analyzed by restriction digestion and DNA sequencing to determine whether they contain an insert of the same size or similar as the mRNA size on a Northern blot.

Once one or more overlapping cDNA clones are identified, the complete sequence of the clones can be determined. The preferred method is to use exonuclease III digestion (McCombie, W.R., Kirkness, E., Fleming, J.T., Kerlavage, A.R., Iovannisci, D.M., and Martin-Gallardo, R., *Methods*: 3: 33-40, 1991). A series of deletion clones is generated, each of which is sequenced. The resulting overlapping sequences are assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a highly accurate final sequence.

A similar screening and clone selection approach can be applied to obtaining cosmid or lambda clones from a genomic DNA library that contains the complete gene from which the EST was derived (Kirkness, E.F., Kusiak, J.W., Menninger, J.,

Gocayne, J.D., Ward, D.C., and Venter, J.C. *Genomics* 10: 985-995 (1991). Although the process is much more laborious, these genomic clones can be sequenced in their entirety also. A shotgun approach is preferred to sequencing clones with inserts longer than 10 kb (genomic cosmid and lambda clones). In shotgun sequencing, the clone is randomly broken into many small pieces, each of which is partially sequenced. The sequence fragments are then aligned to produce the final contiguous sequence with high redundancy. An intermediate approach is to sequence just the promoter region and the intron-exon boundaries and to estimate the size of the introns by restriction endonuclease digestion (ibid.).

Using the sequence information provided herein, the polynucleotides of the present invention can be derived from natural sources or synthesized using known methods. The sequences falling within the scope of the present invention are not limited to the specific sequences described, but include human allelic and species variations thereof and portions thereof of at least 15-18 bases. (Sequences of at least 15-18 bases can be used, for example, as PCR primers or as DNA probes.) In addition, the invention includes the entire coding sequence associated with the specific polynucleotide sequence of bases described in the Sequence Listing, as well as portions of the entire coding sequence of at least 15-18 bases and allelic and species variations thereof. Furthermore, to accommodate codon variability, the invention includes sequences coding for the same amino acid sequences as do the specific sequences disclosed herein. Finally, although the error rate in the automated sequencing used in the present invention is small, there remains some chance of error. Therefore, claims to particular sequences should not be so narrowly construed as to require inclusion of erroneously identified bases or to exclude corrections.

Any specific sequence disclosed herein can be readily screened for errors by resequencing each EST in both directions (i.e., sequence both strands of cDNA).

-19-

The sequences, constructs, vectors, clones, and other materials comprising the present invention can advantageously be in enriched or isolated form. As used herein, "enriched" means that the concentration of the material is at least about 2, 5, 10, 100, or 1000 times its natural concentration (for example), advantageously 0.01%, by weight, preferably at least about 0.1% by weight. Enriched preparations of about 0.5%, 1%, 5%, 10%, and 20% by weight are also contemplated. Further, removal of clones corresponding to ribosomal RNA and "housekeeping" genes and clones without human cDNA inserts results in a library that is "enriched" in the desired clones.

The term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide present in a living animal is not isolated, but the same polynucleotide, separated from some or all of the coexisting materials in the natural system, is isolated.

It is also advantageous that the sequences be in purified form. The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual EST clones isolated from a cDNA library have been conventionally purified to electrophoretic homogeneity. The sequences obtained from these clones could not be obtained directly either from the library or from total human DNA. The cDNA clones are not naturally occurring as such, but rather are obtained via manipulation of a partially purified naturally occurring substance (messenger RNA). The conversion of mRNA into a cDNA library involves the creation of a synthetic substance (cDNA) and pure individual cDNA clones can be isolated from the synthetic library by clonal selection. Thus, creating a cDNA library from messenger RNA and subsequently isolating individual clones from that library results in an approximately 10^6 -fold purification of the native message. Purification of starting material or

-20-

natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly contemplated.

5 In a cDNA library there are many species of mRNA represented. Each cDNA clone can be interesting in its own right, but must be isolated from the library before further experimentation can be completed. In order to sequence any specific cDNA, it must be removed and separated (i.e. isolated and purified) from all the other sequences. This
10 can be accomplished by many techniques known to those of skill in the art. These procedures normally involve identification of a bacterial colony containing the cDNA of interest and further amplification of that bacteria. Once a cDNA is separated from the mixed clone library, it can be
15 used as a template for further procedures such as nucleotide sequencing.

Although claims to large numbers of ESTs and corresponding sequences are presented herein, the invention is not limited to these particular groupings of sequences.
20 Thus, individual sequences are considered as applicants' discoveries or inventions, as are subgroupings of sequences. All of the functional subgroupings set forth in the tables define groupings for which separate claims are contemplated as being within the scope of this invention. Moreover, in
25 addition to claims to individual clones, it is intended that the present disclosure also support claims to numerical subgroupings. Thus, subgroupings of 50 ESTs (and corresponding sequences) are contemplated (e.g., SEQ ID NOS 1-50, 51-100, 101-150, etc.) as being within the scope of
30 this invention, as are subgroupings of 5, 10, 25, 100, 200, and 500 ESTs and corresponding sequences.

III. DNA Constructs

35 The present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs comprise a vector, such as

-21-

a plasmid or viral vector, into which a sequence of the invention has been inserted, in a sense or antisense orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example.

Bacterial: pBs, phagescript, ϕ X174, pBluescript SK, pBs KS, pNH8a, pNH16a, pNH18a, pNH46a (Stratagene); pTrc99A, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia).

Eukaryotic: pWLneo, pSV2cat, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia).

Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, and trc. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In a further embodiment, the present invention relates to host cells containing the above-described construct. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a procaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., *Basic Methods in Molecular Biology*, (1986)).

The constructs in host cells can be used in a conventional manner to produce the gene product coded by the recombinant sequence. Alternatively, the encoded polypeptide

-22-

can be synthetically produced by conventional peptide synthesizers.

Certain ESTs have already been preliminarily categorized by analogy to related sequences in other organisms (see Table 2). Table 10 of Example 10 categorizes particular ESTs broadly as metabolic, regulatory, and structural sequences where known. Constructs comprising genes or coding sequences corresponding to each of these categories are, therefore, specifically and individually contemplated.

Table 11 more particularly separates 127 new ESTs into 13 categories using a different criteria. These are genes related to cell surface; developmental control; energy metabolism; kinase and phosphatase; oncogenes; other metabolism-related polypeptides; peptidases and peptidase inhibitors; receptors; structural and cytoskeletal; signal transduction; transporters; transcription, translation, and subcellular localization; and transcription factors. Table 11 further identifies the EST by the particular gene product for which it apparently codes. Each of these categories individually comprises a preferred category of EST, and preferred constructs and resulting polypeptide can be prepared from those ESTs or the corresponding complete gene sequence.

IV. ESTs and Corresponding Sequences as Reagents

Each of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. The sequences can be used as diagnostic probes for the presence of a specific mRNA in a particular cell type. In addition, these sequences can be used as diagnostic probes suitable for use in genetic linkage analysis (polymorphisms). Further, the sequences can be used as probes for locating gene regions associated with genetic disease, as explained in more detail below.

The EST and complete gene sequences of the present invention are also valuable for chromosome identification.

-23-

Each sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The present invention constitutes a major expansion of available chromosome markers. One hundred ESTs have already been mapped to chromosomes. Using the techniques described in Example 5 or 6, the remaining ESTs and the corresponding complete sequences can similarly be mapped to chromosomes. The mapping of ESTs and cDNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the ESTs. Computer analysis of the ESTs is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the EST will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular EST to a particular chromosome. Three or more clones can be assigned per day using a single thermal cycler. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map an EST to its chromosome include in situ hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific cDNA libraries. Results of mapping ESTs to chromosomal segments are listed in Tables 3 and 4.

Fluorescence in situ hybridization (FISH) of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 500 or 600 bases; however, clones larger than 2,000 bp have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. FISH requires use of the clone from which the EST was derived, and the longer the better. 2,000 bp is good, 4,000 is better, and more than 4,000 is probably not necessary to get good results a reasonable percentage of the time. For a review of this technique, see Verma et al., **Human Chromosomes: a Manual of Basic Techniques**. Pergamon Press, New York (1988).

Reagents for chromosome mapping can be used individually (to mark a single chromosome or a single site on that chromosome) or as panels of reagents (for marking multiple sites and/or multiple chromosomes). Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping (see Tables 8 and 9).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, **Mendelian Inheritance in Man** (available on line through Johns Hopkins University Welch Medical Library).) The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then

-25-

the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb.)

Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that cDNA sequence. Ultimately, complete sequencing of genes from several individuals is required to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

In addition to the foregoing, the sequences of the invention, as broadly described, can be used to control gene expression through triple helix formation or antisense DNA or RNA, both of which methods are based on binding of a polynucleotide sequence to DNA or RNA. Polynucleotides suitable for use in these methods are usually 20 to 40 bases in length and are designed to be complementary to a region of the gene involved in transcription (triple helix - see Lee et al, *Nucl. Acids Res.* 6: 3073 (1979); Cooney et al, *Science* 241: 456 (1988); and Dervan et al, *Science* 251: 1360 (1991)) or to the mRNA itself (antisense - Okano, J. *Neurochem.* 56: 560 (1991); *Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression*, CRC Press, Boca Raton, FL (1988)). Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques have been demonstrated to be efficient in model systems. Information contained in the sequences of the present invention is necessary for the design of an antisense or triple helix oligonucleotide.

-26-

The present invention is also useful tool in gene therapy, which requires isolation of the disease-associated gene in question as a prerequisite to the insertion of a normal gene into an organism to correct a genetic defect.
5 high specificity of the cDNA probes according to this invention have promise of targeting such gene locations in a highly accurate manner.

The sequences of the present invention, as broadly defined, are also useful for identification of individuals
10 from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on
15 a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional
20 DNA markers for RFLP.

However, RFLP is a pattern based technique, which does not directly focus on the actual DNA sequence of the individual. The sequences of the present invention can be used to provide an alternative technique that determines the
25 actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA. One can, for example, take an EST of the invention and prepare two PCR primers from the 5' and 3' ends of the EST.
30 These are used to amplify an individual's DNA, corresponding to the EST. The amplified DNA is sequenced.

Panels of corresponding DNA sequences from individuals, made this way, can provide unique individual identifications, as each individual will have a unique set of such DNA
35 sequences, due to allelic differences. The sequences of the present invention can be used to particular advantage to

-27-

obtain such identification sequences from individuals and from tissue, as explained in Examples 12 - 14.

5 The EST sequences from Examples 1 and 2 and the complete sequences from Example 13 uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the ESTs or
10 complete coding sequences comprising a part of the present invention can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences
15 are necessary to differentiate individuals. The noncoding sequences of Table 9 for example, could comfortably provide positive individual identification with a panel of perhaps 100 to 1,000 primers which each yield a noncoding amplified sequence of 100 bp. If predicted coding sequences, such as
20 those from Table 6, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from ESTs or complete sequences of this invention is used to generate a unique ID database
25 for an individual, those same reagents can later be used to identify tissue from that individual. Positive identification of that individual, living or dead can be made from extremely small tissue samples.

Another use for DNA-based identification techniques is
30 in forensic biology. PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc. In one prior art technique, gene sequences are amplified at specific loci known to contain a
35 large number of allelic variations, for example the DQ α class II HLA gene (Erlich, H., PCR Technology, Freeman and Co.

-28-

(1992)). Once this specific area of the genome is amplified, it is digested with one or more restriction enzymes to yield an identifying set of bands on a Southern blot probed with DNA corresponding to the DQ α class II HLA gene.

5 The sequences of the present invention can be used to provide polynucleotide reagents specifically targeted to additional loci in the human genome, and can enhance the reliability of DNA-based forensic identifications. Those sequences targeted to noncoding regions (see, e.g., Tables 8
10 and 9) are particularly appropriate. As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Reagents for obtaining such sequence information are within the scope of
15 the present invention. Such reagents can comprise complete ESTs or corresponding coding regions, or fragments of either of at least 15 bp, preferably at least 18 bp.

 There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for
20 example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the ESTs or complete sequences of the present invention. Panels of such reagents can identify tissue by species and/or
25 by organ type. In a similar fashion, these reagents can be used to screen tissue culture for contamination.

V. Production of Polypeptide Corresponding to ESTs

 As previously explained, each EST corresponds not only
30 to a coding region, but also to a polypeptide. Once the coding sequence is known, or the gene is cloned which encodes the polypeptide, conventional techniques in molecular biology can be used to obtain the polypeptide.

 At the simplest level, the amino acid sequence encoded
35 by the polynucleotide sequence can be synthesized using commercially available peptide synthesizers. This is

-29-

particularly useful in producing small peptides and fragments of larger polypeptides. (Fragments are useful, for example, in generating antibodies against the native polypeptide.)

Alternatively, the DNA encoding the desired polypeptide
5 can be inserted into a host organism and expressed. The organism can be a bacterium, yeast, cell line, or multicellular plant or animal. The literature is replete with examples of suitable host organisms and expression techniques. For example, naked polynucleotide (DNA or mRNA)
10 can be injected directly into muscle tissue of mammals, where it is expressed. This methodology can be used to deliver the polypeptide to the animal, or to generate an immune response against a foreign polypeptide. Wolff, et al., *Science* 247:1465 (1990); Felgner, et al., *Nature* 349:351 (1991).
15 Alternatively, the coding sequence, together with appropriate regulatory regions (i.e., a construct), can be inserted into a vector, which is then used to transfect a cell. The cell (which may or may not be part of a larger organism) then expresses the polypeptide. (See Example 25.)

20 Antibodies generated against the polypeptide corresponding to a sequence of the present invention can be obtained by direct injection of the naked polypeptide into an animal (as above) or by administering the polypeptide to an animal, preferably a nonhuman. The antibody so obtained will
25 then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies binding the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.
30 Moreover, a panel of such antibodies, specific to a large number of polypeptides, can be used to identify and differentiate such tissue.

VI. Examples

35 Certain aspects of the present invention are described in greater detail in the non-limiting Examples that follow.

-30-

EXAMPLE 1

cDNA Sequences Determined by Random
Clone Selection: First set

5

METHODOLOGY:

With reference to the data presented in Table 1, lambda ZAP libraries were converted en masse to pBluescript plasmids, transfected into E. coli XL1-Blue cells, and plated on X-gal/IPTG/ampicillin plates. A total of 1058 clones were picked at random from three human brain cDNA libraries: fetal brain, two-year-old hippocampus, and two-year-old temporal cortex (Stratagene catalog #936206, 936205, 935, respectively. Stratagene, 11099 N. Torrey Pines Rd., La Jolla, CA 92037). An analysis of these clones is summarized in Table I (see below) In addition, clones selected from the hippocampus library were also analyzed after subtractive hybridization with the fibroblast library. These results are listed in the "Hippocampus Subtracted" column of Table 1.

Templates for DNA sequencing were PCR products or plasmids prepared by the alkaline lysis method. About half of the templates prepared by PCR failed to yield an amplified fragment suitable for sequencing. This was primarily due to use of PCR conditions that minimized the need for further purification of the product but also selected against amplification of long inserts (5 μ l fresh or frozen overnight culture of E. coli carrying the pBluescript plasmid, 7.5 μ M each dNTP, and 0.1 μ M each primer for 35 cycles: 94°C, 40 sec; 55°C, 40 sec; 72°C, 90 sec). A further percentage of the PCR-generated templates failed to sequence, largely due to primer-dimer or other amplification artifacts. Qiagen™ columns improved the percentage of plasmid templates, increasing the yields of usable sequence from about 60% with a standard alkaline lysis protocol to over 90%. Overall, 117 PCR-generated templates and 497 plasmid templates resulted in usable sequence. Dideoxy chain termination sequencing reactions were performed with fluorescent dye-labeled M13

-31-

universal or reverse primers. After a cycle sequencing protocol, carried out in a Perkin-Elmer thermal cycler, sequencing reactions were run on an Applied Biosystems, Inc. (Foster City, CA) 373A automated DNA sequencer. (Cycle sequencing was performed in a Perkin Elmer Thermal Cycler for 15 cycles of 95°C, 30 sec; 60°C, 1 sec; 70°C, 60 sec and 15 cycles of 95°C, 30 sec; 70°C, 60 sec with the Applied Biosystems, Inc. Taq Dye Primer Cycle Sequencing Core Kit protocol). Some sequencing reactions were performed on an ABI robotic workstation (Cathcart, *Nature* 347: 310 (1990) hereby incorporated by reference).

RESULTS:

Singe-run DNA sequence data were obtained from 609 randomly chosen cDNA clones. The number of clones sequenced from each library is summarized in Table 1. Double-stranded cDNA clones in the pBluescript vector were sequenced by a cycle sequencing protocol with dye-labeled primers and Applied Biosystems, Inc. 373A DNA Sequences. The average length of usable sequence was 397 bases with a standard deviation of 99 bases.

Subtractive hybridization has been used successfully to reduce the population of highly represented sequences in a cDNA library by selectively removing sequences shared by another library. (Schmid and Girou, *Neurochem.* 48: 307 (1987); Fagnoli et al, *Anal. Biochem.* 187: 364 (1990); Duguid and Dinauer, *Nucl. Acids. Res.* 18: 2789 (1990); Schweinfest, et al, *Genet. Anal. Techn. Appl.* 7: 64 (1990); Travis and Sutcliffe, *Proc. Natl. Acad. Sci. USA* 85: 1696 (1988); Kato, *Eur. J. Neurosci.* 2: 704 (1990)). Subtractive hybridization was therefore tested as a way of enhancing the number of brain-specific clones in the hippocampus library by hybridizing the hippocampus library with a WI38 human lung fibroblast cell line cDNA library and removing the common sequences (Schweinfest et al, *Genet. Anal. Techn. Appl.* 7: 64 (1990); Sive and St. John, *Nucl. Acids Res.* 16: 10937

-32-

(1988)). Clones from this subtraction are listed in the column "Hippocampus Subtracted" in Table 1.

The EST sequences from this Example 1 are identified as SEQ ID NOS 1-315.

TABLE 1. cDNA Library Composition Determined
By Random Clone Sequencing

-----cDNA Library-----

EST Category	Hippocampus		Hippocampus Subtracted		Fetal Brain		Temporal Cortex	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Databases Match--Human	48	12.8	10	8.6	3	7.9	6	7.5
Mitochondrial Genes	39	10.4	14	12.2	6	15.8	0	0
Repeats: Alu, Line-1, etc.	10	2.7	7	6.0	0	0	11	13.8
Ribosomal RNA	32	8.6	7	6.0	4	10.5	0	0
Other Nuclear Genes	32	8.6	7	6.0	5	13.2	4	5.0
Database Match--Other	160	42.8	44	37.9	20	52.6	6	7.5
No Database Match	53	14.1	24	20.7	0	0	27	33.7
poly A Insert	1	0.3	3	2.6	0	0	26	32.5
No Insert								

EXAMPLE 2

Sequencing of Additional ESTs: Second set

Over 2600 additional cDNA clones have been isolated, partially sequenced and screened. The clones were isolated from four human brain cDNA libraries. The new sequences thus discovered, together with the 315 brain ESTs from Example 1, correspond to over 2400 new human genes. These data represent an approximate doubling of the number of human genes identified by DNA sequencing.

Specifically, four cDNA libraries were used as sources of clones for sequencing. Human hippocampus and fetal brain libraries, plasmid template preparation, sequencing reactions, and automated sequencing were performed as described (Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merrill, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., & Venter, J.C. *Science*, 252: 1651-56 (1991)). A pooled probe consisting of inserts from 10 different EST clones with sequences that matched either mitochondrial genes or the 18S or 28S ribosomal RNAs was used to prescreen a gridded filter array of the hippocampus library; nonhybridizing clones are referred to as the "prescreened library". Another fetal brain library was constructed by and was a gift from Bento Soares (Columbia University). A directionally-cloned library was prepared using the method of Rubenstein, et al. (Rubenstein, J., Elizabeth, A., Brice, A., Ciaranello, R., Denney, D., Porteus, M. & Usdin, T. *Nucl. Acids Res.* 18: 4833-4842) using human adult brain mRNA purchased from Clontech (Palo Alto, CA; Catalogue # 6516-1). Of 482 clones analyzed by restriction enzyme digestion, 33% contained inserts at least 1500 base pairs in length. Stratagene hippocampus and fetal brain library totals include data from Adams et al *Science* 252: 1651.

Sequences of nuclear-encoded cDNAs that did not include interspersed repeats (Schmid, C. W. & Jelinek, W. R. *Science*

-35-

216: 1065-1070 (1982); Paulson, K. E., Deka, N., Schmid, C. W., Misra, R., Schlinder, C. W., Rush, M. G., Kadyk, L., & Leinwand, L. *Nature* 316: 359-361 (1985); Fanning, T. G. & Singer, M. F. *Biochem. Biophys. Acta* 910: 203-212 (1987))
5 were searched against all of GenBank and, in 6-frame translation, against a comprehensive, non-redundant peptide database using the network BLAST (Altschul, S. F., Gish, W., Miller, W., Myers, E.W., & Lipman, D. J. *Mol. Biol.* 215: 403-410 (1990)) server at the National Center for
10 Biotechnology Information. BLAST output was parsed, and an interactive alignment editor was used to select which matches, if any, from each search to record in a relational EST database, which was developed to track sequencing, identification, tissue localization, physical mapping, and
15 the public distribution of the clones, mapping and sequence data. For significant similarities, a putative gene name and Protein Identification Resource (PIR) gene family identification (Barker, W., George, D., Hunt, L., & Garavelli, J. *Nucl. Acids Res.* 19 (Suppl): 2231-2236 (1991))
20 for the EST were assigned. ESTs without significant matches using BLAST were searched in translation against PIR using FASTA. Ten additional marginal matches were found. A total of 2300 new EST sequences comprising 765,505 nucleotides from the current data set have been submitted to GenBank and
25 assigned accession numbers M77851-M79278 and M85308-M86179. All ESTs except those multiply representing actin, tubulin, and myelin basic protein clones were submitted. ATCC accession numbers of cDNA clones from which ESTs were derived are 77501-78999 and 81000-81756. The Genome Data Base
30 expressed D-segment numbers for these clones are D0S1E - D0S2422E. The ESTs from this Example are identified herein as SEQ ID NOs 316-2407.

EXAMPLE 3

EST Characterization: First Set

ESTs including SEQ ID NOS 1-315 were analyzed as follows. Initially, the EST sequences were examined for similarities in the GenBank nucleic acid database (GenBank Release 65.0), Protein Information Resource Release 26.0 (PIR), and ProSite (MacPattern from the EMBL data library, Fuchs R. *Comput. Appl. Biosci.* 7: 105 (1990) Release 5.0 were used). BLAST was used to search Genbank and the PIR (both maintained by the National Center for Biotechnology Information) ESTs without exact GenBank matches were translated in all six reading frames and each translation was compared with the protein sequence database PIR and the ProSite protein motif database. Comparisons with the ProSite motif database were done by means of the program MacPattern from the EMBL Data Library. GenBank and PIR searches were conducted with the "basic local alignment search tool" programs for nucleotide (BLASTN) and peptide (BLASTX) comparisons (Altschul et al, *J. Mol. Biol.* 215: 403 (1990)). PIR searches were run on the National Center for Biotechnology Information BLAST network service. The BLAST programs contain a very rapid database-searching algorithm that searches for local areas of similarity between two sequences and then extends the alignments on the basis of defined match and mismatch criteria. The algorithm does not consider the potential gaps to improve the alignment, thus sacrificing some sensitivity for a 6-80 fold increase in speed over other database-searching programs such as FASTA (Pegarson and Lipman, *Proc. Natl. Acad. Sci. USA*, 85: 2444 (1988)).

Sequence similarities identified by the BLAST programs were considered statistically significant with a Poisson P-value than 0.01. The Poisson P-value less than the probability of as high a score occurring by chance given the number of residues in the query sequence and the database.

-37-

After the BLASTN search, 30 unmatched ESTs were compared against GenBank by FASTA to determine if significant matches were missed due to the use of BLASTN for the database search. No additional statistically significant matches were found. Statistical significance does not necessarily mean functional similarity; some of the reported matches may indicate the presence of a conserved domain or motif or simply a common protein structure pattern. Those ESTs identified as fully corresponding to known human genes or proteins are not included in this disclosure. Statistically significant matches are reported in Table 2, together with the length and percent identity or similarity of each alignment.

On the basis of database searches, 609 EST sequences were classified into eight groups as shown in Table 1 (see Example 1 above). Four groups, with 197 or 32% of the sequences, consist of matches to human sequences: repetitive elements, mitochondrial genes, ribosomal RNA genes, and other nuclear genes. Forty-eight (8%) of the sequences matched non-human entries in GenBank or PIR while 230 (38%) had no significant matches. The remaining 134 (22%) sequences contained no insert or consisted entirely of polyA between the EcoRI cloning sites.

Thirty-six ESTs matched previously sequenced human nuclear genes with more than 97% identity. Four of these ESTs are from genes encoding enzymes involved in maintaining metabolic energy, including ADP/ATP translocase, aldolase C, hexokinase, and phosphoglycerate kinase. Human homologs of genes for the bovine mitochondrial ATP synthase $F_0\beta$ -subunit and porcine aconitase were also found (Table 2). Brain-specific cDNAs included synaptophysin, glial fibrillary acidic protein (GFAP), and neurofilament light chain. At least six ESTs are from genes encoding proteins involved in signal transduction: 2',3'-cyclic nucleotide 3'-phosphodiesterase (2 ESTs), calmodulin, c-erbA- α -2, $G_s\alpha$, and Na^+/K^+ ATPase α -subunit. Other ESTs were matches to genes for ubiquitous structural proteins -- actins, tubulins, and

-38-

fodrin (non-erythroid spectrin). ESTs also document the presence in the hippocampus cDNA library of the ret proto-oncogene, the ras-related gene rhoB, and one of the chromosome 22 breakpoint cluster region transcripts. Eight
5 ESTs are from genes known to be associated with genetic disorders (Online Mendelian Inheritance in Man). More than half of the human-matched ESTs from Example 1 have been mapped to chromosomes, indicating the bias of GenBank entries toward well-studied genes and proteins.

10 ESTs without significant GenBank matches were also compared to the ProSite database of recognized protein motifs. Not counting post-translational-modification signatures, fifty-four sequences contained motifs from the database. Some patterns, particularly the "leucine zipper",
15 are found in scores or hundreds of proteins that do not share the functional property implied by the presence of the motif.

Similarities to sequences from other organisms were also detected in the BLAST searches of GenBank and PIR (Table 2). Several ESTs displayed similarity to "housekeeping" genes,
20 including the ribosomal proteins S10 and L30 (rat) and the above glycolytic enzymes. EST00257 (SEQ ID NO:77) shows strong nucleotide sequence similarity to the squid (67%) and Drosophila (70.4%) kinesin heavy chain. Kinesin was first described as a microtubule-associated motor protein involved
25 in organelle transport in the squid giant axon (Vale et al, Cell 42: 39 (1985)). Six oncogene-related sequences were also among the cDNA clones sequenced. EST00299 (SEQ ID NO:180) and EST00283 (SEQ ID NO:271) show similarity to several ras-related genes and EST00248 (SEQ ID NO:102)
30 matched the 3' untranslated region of the bovine substrate of botulinum toxin ADP-ribosyltransferase. Similarities with an S. cerevisiae RNA polymerase subunit and Torpedo electromotor neuron-associated protein were also observed. Two ESTs may represent new members of known human gene families: EST00270
35 matched the three β -tubulin genes with 88-91% identity and

-39-

EST00271 (SEQ ID NO:248) matched α -actinin with 85% identity at the nucleotide level.

Among the most interesting of the primary sequence relationships was the similarity of ESTs to the *Drosophila* genes Notch and Enhancer of split. Nucleotide and peptide alignments of EST00256 (SEQ ID NO:188) and EST00259 (SEQ ID NO:227) with the *Drosophila* genes have been demonstrated. Both genes are part of a signal cascade encoded by the "neurogenic" genes that are involved in the differentiation of neuronal and epidermal cell lineages in the neuroectoderm of the developing *Drosophila* embryo (Campos-Ortega, *Trends in Neuro. Sci.* 11: 400 (1988)). It has been proposed that the Enhancer of split protein interacts with a membrane protein that is the product of the Notch gene to convert a developmental signal into an altered pattern of gene expression (id. *J. Mol. Biol.* 215: 403 (1990)). EST00256 (SEQ ID NO:188) matches near the 5' end of the Enhancer of split coding sequence, away from the mammalian G protein β subunit- and yeast cdc4-like elements (Hartley et al, *Cell* 55: 785 (1988); Klambt et al. *EMBO J.* 8: 203 (1989)). Part of the EST00259 (SEQ ID NO:227) match to Notch in the cdc10/SW16 region that is similar to three cell-cycle control genes in yeast and is tightly conserved in the *Xenopus* Notch homolog, Xotch. In *Drosophila*, Enhancer of split is absolutely required for formation of epidermal tissue. Notch contains several epidermal growth factor-like repeats and appears to play a general role in cell-cell communication during development (Banerjee and Zipursky, *Neuron* 4:177 (1990)).

Seven genes were represented by more than one EST. Comparisons of all the ESTs against one another revealed two overlaps of unknown ESTs: EST00233 (SEQ ID NO:32) and EST00234 (SEQ ID NO:8) match in opposite orientations and EST00235 (SEQ ID NO:204) and EST00236 (SEQ ID NO:148) match in the same orientation beginning at the same nucleotide. Five human genes were represented by more than one EST: β -

actin (3), λ -actin (2), α -tubulin (2), α -2-macroglobulin (2), and 2'3'-cyclic-nucleotide-3'-phosphodiesterase (2). Those few instances where two or more ESTs represent different portions of a single cDNA can be readily ascertained when the sequence of the full cDNA insert is determined in accordance with Example 13.

Example 4

EST Sequences Characterization: Second Set

The ESTs of Example 2, including SEQ ID NOs 316-2407, were screened against known sequences listed in GenBank and other databases, as in Example 3. The results are reported in Table 2. The quality of the match is given as percent identity and length in base pairs for nucleotide matches and amino acid residues for peptide matches. In many cases ESTs match multiple domains on several related proteins; for example, EST00825 matches two transmembrane domains on both GABA and Norepinephrine transporters. Nucleotide databases are: GenBank (GB), and EMBL (E); peptide databases are: GenPept (GPU), Swiss-Prot (SP), and PIR.

The great majority (83%) of the partial cDNA sequences reported in Example 2 are unrelated to any sequences previously described in the literature. Based on database matches to known genes from humans as well as from such evolutionarily distant organisms as *E. coli*, yeast, *C. elegans*, *Drosophila*, barley, *Arabidopsis*, rice, and green algae, we have preliminarily identified the functional type of a number of the ESTs (Table 2). These include a novel gene similar to Notch/Tan-1 (Adams et al., *supra*), a new neurotransmitter transporter gene, and a new member of the multi-drug resistance gene family. Several genes involved in development or cell differentiation in *Drosophila* are represented by similar human ESTs, including seven in *absentia* (Carthew, R. & Rubin, G. Cell 63: 561-577 (1990)),

-41-

big-brain (bib) (Rao, Y., Jan, L., & Jan, Y. *Nature* 345: 163-167 (1990)), the discs tumor suppressor (Woods, D. & Bryant, P. *Cell* 66: 1-20 (1991)), and the homeotic gene orthodenticle (Finkelstein, R., Smouse, D. Capaci, T., Spradling, A. & Perrimon, N. *Genes. Dev.* 4: 1516-1527 (1990)). New members of gene families previously known in humans include a Ca^{+2} -transporting ATPase, an ADP ribosylation factor, and a new neural-cell adhesion molecule gene.

The 1971 ESTs without a putative identification were analyzed using the coding-region prediction program CRM via the GRAIL server (Uberbacher, E. & Mural, R. *Proc. Natl. Acad. Sci. USA* 88: 11261-5 (1991)). Fifteen percent of the unknown ESTs scored an excellent probability of containing protein-coding sequence. Fifty percent of the ESTs to known human genes contain protein-coding sequences, therefore, at most half of the unknown ESTs are likely to contain coding sequences. We have found no evidence that genomic DNA or cDNA to unspliced precursor RNA is a major contaminant of either the hippocampus or fetal brain library.

-42-

Table 2: ESTs Identified by Database Matches

SEQ ID	EST#	Putative Identification	Accession	DB	Len	%ID
208	EST00250	60K filarial antigen	A28209	PIR	108	56.9
2320	EST01784	60K filarial antigen	A28209	PIR	88	50.6
969	EST01982	ADP-ribosylation factor 1	B33283	PIR	84	41.2
1834	EST01620	AMP deaminase, brain	A37056	PIR	57	100.0
97	EST00289	Aconitase	A35544	PIR	105	90.6
251	EST00370	Actin, other	S10021	PIR	44	51.1
248	EST00271	Actinin, alpha	HUMACTAR	GB	271	85.3
891	EST01891	Actinin, alpha	HUMACTAR	GB	315	81.6
1500	EST02538	Actinin, alpha	HUMACTAR	GB	271	75.0
132	EST00110	Agrin	RATAGR	GB	269	82.2
1852	EST01825	Agrin	RATAGR	GB	103	84.6
1094	EST02113	Ala	HUMALA	GB	92	82.8
691	EST00675	Alcohol dehydrogenase	RICGOS2G_1	GPU	38	59.0
2408	EST00244	Amyloid A4	HUMAFPA4	GB	135	91.9
1965	EST01664	Amyloid A4	A29030	PIR	52	54.7
2068	EST01694	Amyloid A4	QRHUA4	PIR	83	69.0
2092	EST01700	Anion exchanger homolog AE3	A33638	PIR	95	97.9
1880	EST01634	Axonal glycoprotein TAG-1	A34695	PIR	69	87.1
1492	EST02530	B cell-specific Mo-MLV integration site 1 (bmi-1)	MUSBMI1A	GB	111	87.5
1277	EST02306	Bib protein	S09699	PIR	57	53.4
13	EST00255	Cadherins	CADN\$HUMAN	SP	41	45.2
1348	EST02378	cAMP-dependent protein kinase inhibitor	MUSPKI	GB	234	91.5
1931	EST01041	cAMP-regulated phosphoprotein	B35308	PIR	21	86.4
1413	EST02447	cAMP-specific phosphodiesterase	HUMPDEAA	GB	363	69.0
396	EST01443	CDPdiacylglycerol-serine O-phosphatidyltransferase	JH0368	PIR	33	41.2
1956	EST01663	Ca2+-transporting ATPase 2	B28085	PIR	125	88.9
1126	EST02146	Calbindin D28	RATCALBD28	GB	81	87.8
1039	EST02055	Calcium channel	S05054	PIR	33	67.6
1910	EST01645	Calmodulin	RATRCM1	GB	120	90.1
485	EST01466	Calmodulin-dependent protein kinase, type II, beta	A26464	PIR	93	98.9
913	EST01913	Clathrin coat assembly protein AP50 homolog	YSCYAP54_1	GPU	62	63.5
2004	EST01676	Cofilin	PIGCOFIL	GB	132	89.5
2400	EST01824	Cysteine-rich intestinal protein	GYRTI	PIR	56	66.7
1588	EST02633	D2223 repetitive DNA	HUMREP	GB	160	76.4
2192	EST01257	Diacylglycerol kinase, lymphocyte	S09156	PIR	44	42.2
1441	EST02477	Diamine acetyltransferase	ATDA\$HUMAN	SP	74	45.3
650	EST00642	Dilute (myosin heavy chain)	MUSDILUTE_1	GPU	27	100.0
2302	EST01779	Disca-large tumor suppressor	DRODLGA_1	GPU	53	83.0
188	EST00256	Enhancer of split	A30047	PIR	86	58.6
2289	EST01325	Fatty acid synthase	RATFAS	GB	98	79.8
310	EST00377	Fo ATPase beta subunit, mitochondrial	BOVMTASB	GB	293	85.4
1332	EST02362	GA binding protein, beta subunit	MUSGAC_1	GPU	86	90.8
1667	EST00825	Gamma-aminobutyric acid transporter	A35918	PIR	26	59.3
2217	EST01738	Gelation factor ABP-280	A37098	PIR	74	80.0
1412	EST02446	Glutamate-aspartate carrier protein	JV0092	PIR	57	37.9
1020	EST02034	Glutaminase	GLS\$RAT	SP	34	74.3
1885	EST01639	Histocompatibility antigen modifier 1	A37779	PIR	63	75.0
1495	EST02533	Hypothetical 43.5K protein	JU0319	PIR	43	52.3
2326	EST01791	Inositol-1,4,5-trisphosphate 3-kinase	JN0129	PIR	65	68.2
SEQ ID	EST#	Putative Identification	Accession	DB	Len	%ID

SUBSTITUTE SHEET

-43-

724	EST01529	Interferon-induced 54K protein	IN14#HUMAN	SP	76	70.1
1035	EST02051	J1 protein	MUSJ1PRO	GB	362	85.7
1229	EST02258	KUP protein	HUMKUPMR_1	GPU	54	36.4
993	EST02007	Kinase 5 protein	CHKCEK5_1	GPU	68	94.2
77	EST00257	Kinesin	A35075	PIR	57	86.2
78	EST00258	Kinesin	A35075	PIR	62	47.6
2245	EST01748	Kinesin	A35075	PIR	98	52.5
2282	EST01764	Lamin B receptor	A38427	PIR	76	71.4
2173	EST01724	Lon protease	JQ0901	PIR	103	41.3
1427	EST02463	Long-chain-fatty-acid-CoA ligase	A36275	PIR	36	62.2
313	EST00276	Lysosomal membrane glycoprotein 1 (LAMP-1)	A31959	PIR	53	48.3
181	EST00247	MARCKS (myristoylated alanine-rich protein kinase	BOVMARCKS	GB	139	83.6
1386	EST02418	MARCKS homolog	MMF52	EU	237	92.4
769	EST00734	MARCKS homolog	S08341	PIR	61	40.3
43	EST00371	Maternal G10 protein	S05955	PIR	38	92.3
1468	EST02505	Matrin 3	RATMATRIN3	GB	137	93.5
639	EST00632	Membrane transport superfamily (GTP-dependent)	A24400	PIR	63	39.1
1894	EST01643	Membrane transport superfamily (GTP-dependent)	A24400	PIR	71	50.0
824	EST01865	Microtubule-associated protein 1B	RATNEU	GB	293	86.4
223	EST00368	Microtubule-associated protein 1B	A33845	PIR	30	54.8
2032	EST01683	Microtubule-associated protein 1B	A33845	PIR	49	62.0
2017	EST01678	Milk fat globule membrane protein	A36479	PIR	48	81.2
1704	EST01580	Myeloid differentiation primary response gene MyD1	MUSMYD118_1	GPU	76	88.3
2226	EST01744	NAD(P)+ transhydrogenase (B-specific)	DEBOXM	PIR	88	93.1
1567	EST02610	Neural cell adhesion molecule L1	S05479	PIR	82	43.4
506	EST01471	Neuraxin	S06017	PIR	120	84.3
1566	EST02609	Neutrophil oxidase factor	A34855	PIR	43	47.7
952	EST01981	Notch/Xotch	HUMTAN1_1	GPU	85	57.0
227	EST00259	Notch/Xotch	A35844	PIR	74	85.3
1395	EST02429	Nuclear factor 1-like protein (NF1)	HAMNF1A	GB	111	92.0
1681	EST01573	Nucleoside diphosphate kinase	A33386	PIR	71	52.8
346	EST01828	Otd homeotic protein	A35912	PIR	35	52.8
2254	EST01751	Phosphatidylinositol-4,5-bisphosphate phosphodiester	A28807	PIR	40	90.2
1869	EST00992	Polymyxin B resistance	A32714	PIR	20	76.2
93	EST00287	Processing enhancing protein	S03988	PIR	98	58.8
2353	EST01806	Prohibitin	RATPROHIB_1	GPU	120	97.5
2287	EST01775	Prohormone cleavage enzyme	MUSMPC1A_1	GPU	91	93.5
9	EST00376	Prolyl endopeptidase	PIGPREP	GB	223	83.9
1069	EST02087	Protein kinase C, zeta	HUMPKCL	GB	382	58.7
1933	EST01650	Protein phosphatase 2A beta subunit	HUMPROP2AB	GB	288	76.8
202	EST00298	Protein-tyrosine phosphatase LRP	LRP#MOUSE	SP	62	44.4
1654	EST01572	Protochlorophyllide reductase	S04783	PIR	34	57.1
38	EST00374	RNA polymerase II 6th subunit (RPO26)	A36352	PIR	72	75.3
1478	EST02515	Rab5	F34323	PIR	91	82.6
2368	EST01389	Radial spoke protein 3	S05962	PIR	58	52.5
37	EST00038	ras p21-like small GTP-binding protein (smg GDS)	BOVSMGGDS	GB	131	89.4
180	EST00299	ras-related proteins	S10493	PIR	51	46.1
1700	EST01579	Retrovirus-related gag polyprotein	FOHUE2	PIR	95	77.1
1511	EST02550	Retrovirus-related pol polyprotein	GNLJGL	PIR	50	54.9
102	EST00248	rho H12/ ARH12	BOVBGBRH	GB	195	79.6
1715	EST01583	Ribosomal protein L18a	R5RT18	PIR	68	95.7
SEQ ID	EST#	Putative Identification	Accession	DB	Len	%ID

-44-

1856	EST01627	Ribosomal protein L1a	A24579	PIR	75	63.1
1974	EST01667	Ribosomal protein L3	JQ0771	PIR	74	80.0
301	EST00300	Ribosomal protein L30	R6RT30	PIR	57	96.5
22	EST00301	Ribosomal protein S10	R3RT10	PIR	66	97.0
2402	EST01826	Ribosomal protein S10	R3YM10	PIR	36	51.4
463	EST01459	Ribosomal protein YL10	S11581	PIR	40	68.3
1408	EST02442	Seven in absentia	A36195	PIR	46	80.8
299	EST00249	smg p25A GDP dissociation inhibitor	A35652	PIR	97	77.5
951	EST01960	Spectrin, beta	HUMSPTB	GB	268	67.7
2089	EST01699	Sperm membrane protein	A35981	PIR	52	58.5
2073	EST01697	Succinate dehydrogenase flavoprotein	BOVSDHFP1_1	GPU	44	100.0
2138	EST01715	Succinate dehydrogenase flavoprotein	BOVSDHFP1_1	GPU	49	92.0
430	EST00472	Synaptotagmin (p65)	SY654HUMAN	SP	27	53.6
1371	EST02402	Talin	MUSTALINR_1	GPU	79	81.2
1771	EST01601	Thiosulfate sulfurtransferase (rhodanese)	ROBO	PIR	65	81.8
300	EST00232	Transforming protein (dbl)	TVHUDB	PIR	25	65.4
189	EST00282	trkB	A35104	PIR	33	67.6
653	EST01512	Tubulin, alpha	HUMTUBAG	GB	223	75.0
594	EST01490	Tubulin, beta	HUMT885	GB	298	93.6
757	EST01542	Tubulin, beta	HUMT88M	GB	217	90.4
1245	EST02274	Tubulin, beta	A26581	PIR	105	88.7
1147	EST02169	Tyrosine kinase	HUMECK	GB	384	74.3
1701	EST00863	Unc-104	JN0114	NR	36	45.0
2121	EST01711	Valine-tRNA ligase	A29871	PIR	56	57.9
187	EST00152	Wilm's tumor-related protein	HUMQM	GB	228	99.6
1728	EST01588	XPR2 alkaline extracellular protease	B26955	PIR	88	46.1
249	EST00275	Zinc Finger Proteins	S06551	PIR	25	57.7
413	EST01446	Zinc Finger Proteins	S00754	PIR	45	60.9
469	EST01460	Zinc Finger Proteins	C32891	PIR	34	54.3
833	EST01560	Zinc Finger Proteins	S00754	PIR	105	67.0
1230	EST02259	Zinc finger proteins	S00754	PIR	71	62.5
1496	EST02534	Zinc finger proteins	A34612	PIR	50	45.1
2324	EST01352	Zinc Finger Proteins	S10397	PIR	29	56.7

SUBSTITUTE SHEET

-45-

There is little redundancy in EST sequencing according to the present invention. Of the nuclear-encoded messenger RNAs, the most common ESTs were to the β -actin (0.6% of the EST clones) and myelin basic protein genes (MBP, 0.5% of the clones). MBP, a highly expressed structural component of nerve tissue (Kamholtz, J., de Ferra, F., Puckett, C., & Lazzarini, R. *Proc. Natl. Acad. Sci., USA* 83: 4962-4966 (1986)), displays four alternate splicing forms, of which at least two are present among the ESTs reported here. Other common ESTs were Gs-alpha gamma-actin and both α - and α -tubulin.

By matching ESTs to known database sequences, a phenotypic characterization of the tissue begins to emerge. Protein superfamilies matched by ESTs were grouped into three broad functional categories to assess the biological spectrum represented by these randomly selected cDNA clones. Structural and metabolic classes comprised about 30% of the ESTs with database matches. Twenty-five percent were involved in regulatory pathways and the remainder were not classifiable. Eleven of the eighteen enzymes of glycolysis and the citric acid cycle are represented by at least one subunit or isozyme. In addition, several genes not previously known to be expressed in the brain were matched, including spermine/spermidine acetyltransferase (Casero, R., Celano, P., Ervin, S., Applegren, N., Wiest, L. & Pegg, A. *J. Biol. Chem.* 266: 810-814 (1991)) and osteopontin (Young, M., Kerr, J., Termine, J., Wewer, U., Wang, M., McBride, W. & Fisher, L. *Genomics* 7:491-502 (1990)).

EXAMPLE 5

Mapping of ESTs to Human Chromosomes

Randomly selected ESTs corresponding to SEQ ID NOs. were assigned to chromosomes via PCR (see Table 3). Oligonucleotide primer pairs were designed from EST

-46-

sequences to minimize the chance of amplifying through an intron. The oligonucleotides were 18-23 bp in length and designed for PCR amplification using the computer program INTRON (National Institutes of Mental Health, Bethesda, MD). The program is based on the assumptions that: 1) introns are genomic sequences that interrupt the coding and noncoding sequences of genes (Smith, J. Mol. Evol. 27:45-55 (1988)); 2) there are consensus sequences for splice junctions (Shapiro, et al., Nucl. Acids Res. 15:7155-7174 (1987)); and 3) that 90% of the human genes studied have 3' untranslated regions of mRNA not interrupted by introns in the genomic DNA (Hawkins, Nucl. Acids Res. 16:9893-9908 (1988)).

The program evaluates the likelihood that a given GG or CC dinucleotide represents a former exon-intron boundary. Specifically, every input strand is processed by the INTRON program twice, first evaluating the sense mRNA strand, and then processing the complementary or anti-sense strand. The program evaluates each sequence by finding all GG or CC pairs (possible former splice sites), searching for STOP codons in all three reading frames, and analyzing the GG or CC pairs surrounded by stop codons. All regions of the EST that are unlikely to contain splice junctions based on CC content, GG content, and stop codon frequency are then marked by the program in uppercase.

The creation of PCR primers from known sequences is well known to those with skill in the art. For a review of PCR technology see Erlich, H.A., PCR Technology; Principles and Applications for DNA Amplification. 1992. W.H. Freeman and Co., New York. ESTs were examined for the presence of stop codons in each reading frame and for consensus splice junctions. The presence of stop codons and absence of splice junction sequences are more characteristic of 3' untranslated sequences than of introns. The untranslated sequences are unique to a given gene; thus, primers from

-47-

these regions are less likely to prime other members of a gene family or pseudogenes.

The primers were used in polymerase chain reactions (PCR) to amplify templates from total human genomic DNA. PCR conditions were as follows: 60 ng of genomic DNA was used as a template for PCR with 80 ng of each oligonucleotide primer, 0.6 unit of Tag polymerase, and 1 uCi of a ³²P-labeled deoxycytidine triphosphate. The PCR was performed in a microplate thermocycler (Techne) under the following conditions: 30 cycles of 94°C, 1.4 min; 55°C, 2 min; and 72°C, 2 min; with a final extension at 72°C for 10 min. The amplified products were analyzed on a 6% polyacrylamide sequencing gel and visualized by autoradiography. If the size of the resulting product was equivalent to the EST from which the primers are derived, then the PCR reaction was repeated with DNA templates from two panels of human-rodent somatic cell hybrids; BIOS PCRable DNA (BIOS Corporation) and NIGMS Human-Rodent Somatic Cell Hybrid Mapping Panel Number 1 (NIGMS, Camden, NJ).

PCR was used to screen a series of somatic cell hybrid cell lines containing defined sets of human chromosomes for the presence of a given EST. DNA was isolated from the somatic hybrids and used as starting templates for PCR reactions using the primer pairs from EST sequences selected above. Only those somatic cell hybrids with chromosomes containing the human gene corresponding to the EST will yield an amplified fragment. ESTs were assigned to a chromosome by analysis of the segregation pattern of PCR products from hybrid DNA templates. For a review of techniques and analysis of results from somatic cell gene mapping experiments. (See Ledbetter et al., *Genomics* 6:475-481 (1990).) The single human chromosome present in all cell hybrids that give rise to an amplified fragment represents the chromosome containing that EST.

The assignment of 100 ESTs and corresponding genes to chromosomes by PCR is shown in Table 3.

-49-

Table 3: Assignment of ESTs to Chromosomes by PCR

<u>SEQ ID</u>	<u>EST#</u>	<u>Chr</u>	<u>PRIMER #1</u>	<u>PRIMER #2</u>
5	EST00012	1	TCCAGGCAATCCCAGAATAG	CTAATTGAGCTCACTGGCCC
57	EST00058	1	CTGTTTGCAAGTTTCAAAGC	GCCATTTCTAACAACCAGAG
64	EST00066	1	GCCATTGTGCTGAATAGAGT	GTTAGTGTTCCTTAGCAAG
83	EST00079	1	CAGCTAATTGACCTGGGCTA	CAACATGCTCTGAGCTTTAG
83	EST00079	1	GGCAGAGCATAATGAGTATA	CATATGCATATGGTCCCTAT
91	EST00086	1	AGTTTAGATGGAGGGCTGTC	TCTGCCCTAATGCGCAGGCT
105	EST00365	1	CTTAATCACCTCCCTTTTGT	CCTTAGTTGGAGATAAGGTC
109	EST00095	1	AGTCTAATCCTGTACACTTG	CGGGCTTTCTCTGAATTGGT
116	EST00100	1	TTAGAAGTGCCCATGGGAGG	TTTTAAGGCTCTGGAGTGTT
141	EST00118	1	CTCAGAGAAACTTAGGTGAA	CTACAGAATCATTTCACCAG
220	EST00372	1	AAGTTGCACATTGCCCAAGG	ATAGTACTGCAAGGTTATTTC
237	EST00187	1	TTACAAATTTCTTTGACGC	CTGAAGGAGCACAGTTTCTC
242	EST00192	1	GGATCAGATAATCAAACAGG	GCTTAGGATATGAATGCATA
259	EST00202	1	GCATCACAGTTTAACTGAGG	CTACATATTTGTGCCTCCTT
269	EST00293	1	CTGTTGCTGTGTCAGTAGCTT	CTTTTGACCCAGTGAAACTT
299	EST00249	1	GATCATGCAGACGTAGATAT	CCAACCTCTGCCAGATCATT
1651	EST00810	1	TAGTCGCTGTAAGTTGATTC	GCTTTGCTGGATGCTTCATT
16	EST00021	2	CAGGCAAGTTTCTTCCAGGA	TCAGACCCATGGTCAGCTT
1898	EST01013	2	GGCTGAGAACGGTTAGCATA	CCCTCAGCTTAGGGGAATG
8	EST00234	2	TAGAAGGCCAACTATGTCCC	GGTTGAGGATTGGCTTTTAC
36	EST00037	2	AGCCAGAAGGCTGCTTAAAG	GCAGTGAACCCAGTACTCCTA
123	EST00106	2	GTCTAATTTGTAACTTCAG	GATAGATTGTATAAGAAGCC
192	EST00155	2	GATTTATGTCTGGGAACATA	GCAGCATGTGAAAGAATGAT
200	EST00162	2	TTTAATGGGTGGTGGGAGCT	CGATGCACATCCTTCTCCAT
284	EST00216	2	CCTAAGAATTCGTTTGGCTC	GTCTGGCACATAATAGATTG
102	EST00248	3	ATACTACATCTAGTCTGG	TTACAGTTCTGTGGTTTTT
167	EST00138	3	AAACAGCTGCGGAGTACA	AAAGGATCTCCACTCCAGA
12	EST00274	3	CCTAGCAAACCTCATAACAC	CATAAGTGAATGGACACAGG
60	EST00062	3	ACACATTAACGGGTGCTGCAG	GGAATCAGCCCTTGAGGACT
77	EST00257	3	AAGCTCACAAACGAGATCTG	CTGGAACAGCTTACAAAGGT
107	EST00093	3	ATTGAACTCTGTCAACAGTG	TGTAAACCAAAGGCCAAACT
108	EST00094	3	AL2 - GCAGGATGTCAGTCTTTGAG	AGCACACATTATCTACCACGGC
1706	EST00857	3	AL2 - GCAGGATGTCAGTCTTTGAG	CCAGCACACATTATCTACCACG
37	EST00038	4	AACTTCGCGAGTCATGAGAAC	TGTATCGGGCAGTTCTCAG
6	EST00013	4	CACATGTTCTCCCTCTTTCA	GCATTTTGAGCTCTTCCGT
37	EST00038	4	AL2 - GGAAGTACAGGATTGGC	TTAGAGATGGGATGATGCCG
31	EST00033	5	TGGGTACCCCTAAGGTGTTG	GACTAATCTAAGGTCTAGG
28	EST00030	5	AGATAAGTTAGGAAGCTGGT	ACTCACTGCTAGTATCATCC
59	EST00061	5	AAAGTTTCTTAGCACCCCCC	CAGACTTTGACAAAAGAATC
74	EST00073	5	ATCAGACACGTGGCAGGGTT	AAGTCCCTGAGGGTGCAGAA
121	EST00104	5	TGAAGGCAGCTGCTAAATCT	GGATGTATTGATCTGACTCA
149	EST00123	5	ATACTGTCAACGGAGGGTGA	GTCTGCAGGTTTCTCCTTGA
235	EST00185	5	TTACTGTCCCATCAGATATC	TACACTCTTAAAGAGGTATG
1643	EST00803	5	GAGCGTTTAAAAGAGATTCT	TACAGACAGCCATGTTCCAA
1677	EST00835	5	AL2 - TCTCCAACACAGTCATGC	CGGATGCCATCATATACC
23	EST00026	5	CCTGCAGTGACACTTAACAT	CTGCTCACCTGAAATTGATAC
121	EST00104	5	AL2 - CAGATCAATACATCCTCTGGG	CTGTGCAGTGGTGAGTAAAGG

SUBSTITUTE SHEET

-50-

SEQ ID	EST#	Chr	PRIMER #1	PRIMER #2
1	EST00007	6	TAGTTGATGGTCTGGGTTAT	GAAATCCCAGGGAGACAATG
19	EST00023	6	CAACTTACATTAGGGGTTTG	GACCTCATTAGAAGAGCCCA
155	EST00129	6	GGAAGCTGCCATATAAGCTC	TCAGTGTCTGACAATCTACC
224	EST00356	6	GCTGTATGTTAACCCTTTGT	TGGAACCTCAAACACTGCT
288	EST00219	6	ACTTTTCATGTTGAGAAGTAT	ATCTAGCTGAAACATTGCTG
1638	EST00798	6	CTTCATCTGTTAACTGTTGA	TGAAAATGAGTCACAGGCAG
1675	EST00833	6	AL2 - ACCCAGTTCTCAAAGACC	GGTTTACCATTTCAGAGGC
22	EST00301	6	CTCCGTGATTACCTTCATCT	TTGTAGGTATCTCTGTACGCT
207	EST00167	7	GGTGCTACTTTGTGAATGCT	AGCAATGTGATTTTGTAGG
137	EST00272	7	AGTGGTCACTATCTACATGG	GATTCAGAATTACTAAGCCG
1659	EST00817	7	TGTATAGGCTCTACATAAAG	CTTAATCATGGATTCTTCGT
1680	EST00838	7	AL2 - GTTCTTTCCACGGTATGC	TTGTTGGTACTGAGGAAGTGCG
292	EST00223	8	TGCAGCAGTGACCATGAGAA	ATCATCTTTCCACGCGGCTT
134	EST00375	9	GTTGGGCTTCTGTGTTCAA	CTGGCTGCTCAGCAACTCAT
1906	EST01021	9	GGATGTTTTCTATGTGACGA	TTCCAGTGCCCCCTTTTGTCC
1645	EST00804	10	CTCCTTTGGGACAAACAACT	CCAAACCAACATATTCTA
20	EST00024	10	AGCTGTTCTCTGAGAGATGCA	CCTTGTGAAGAAAGACTTTC
157	EST00131	10	TACGCAACAGGTCACTTGG	CTAAGCATCTGCATGCTCAG
172	EST00142	10	TACTAGCATTTCTTACTCTC	TATGCTGATTGTTTGCACCTC
250	EST00197	10	GGTGATTAGAGAGTCTGTTG	GAACCTCTGTAGTGTTCTAAA
133	EST00111	11	GGAAATTAGGCTTAGCTCAC	GTGCAGAATACTTAGAGTCC
178	EST00294	11	GTTTGAAGGAAGTGATTTCC	TAGGGCCACCTCCAGTTTCAT
10	EST00016	11	GTCCTTTGGATTCTACGTAGA	CGATAATGACATTTCTTCTGG
126	EST00109	11	AL2 - CTAACCACAACCACACATTG	CCTCAGCACAAGAGAAGATGG
7	EST00014	12	AACCTGCAACATAAATACTAG	GAGCAATGATTTCTAACAGT
254	EST00200	13	TTGTGTACTGTCTGTATAGAC	TAAGCCATGGGCATCTATAA
2409	EST00273	13	GCAAGATGATGGAACATCCC	TTCTTCTGAGGGCTCTACA
170	EST00295	14	GGTGCTTAAGGCCACTTTTG	CTTAGAGGATCATAGGTCTG
255	EST00201	14	CCAGGAGAGTAAGAAGATCA	GCAGAGTTGAATATGAACCT
290	EST00221	14	GTGCCAAGATGGCTCATGTA	GTATAGCTTTAAGCCAGTTC
293	EST00224	14	AATGCATTATGCCTGGTCTT	GGAAAAGTCTAGAACTTAGT
1664	EST00822	14	GGGTCAGAATTAAGAGGTCT	GTTTCATCTCTAACTCCTTTC
315	EST00008	14	AAGCTGGCTGGGAAATGTTT	GTCTAGCTAGTAACTTACAC
1689	EST00845	14	AL2 - AGGAGGAAGCTGAAATCC	GGAAAGTCCATAAGAGACTCACC
95	EST00088	15	GTGACAGACCATGTCTATTG	AAGTGAGCGATTGCACCTTC
205	EST00165	15	AGGATGACCTGAGTGAGCTG	CCATGGCAGCAAGGAACTCT
33	EST00034	16	TGTGTGAAAGGGAGTCTTGT	CCATTTTGAAGTGTCCATAG
247	EST00279	16	TGGCTAGGGCAGGCCTTAAA	GAGAAGAATATCAAATGGGG
18	EST00373	16	CCATCTGTGTCCCAATTAAGC	AGGGAAGAAGTCTAGAGCGA
68	EST00068	17	CAAAGACGGGAGACGAATGA	AGTGAACGCGTGCCCTATG
1652	EST00811	17	GAGCTGCATGTTGATAAGTA	TTGACTTAAGCTGACCTTAA
1702	EST00854	17	AL2 - TTGCTGTGGAATCCATGAGAG	GGCAAGTGATCTGTTCTTGG
84	EST00080	19	AGAGATGTCAAGTCCATTATC	CTATTCCACCTTACTCAAGG
223	EST00368	19	CATCATGTCTGGAGACGCATT	TGGATGACCTGAGTCTGCAG
21	EST00025	20	AGTTCCTGGAGGCTAGGAGTT	ATGTAAGGACCCCTAGATGG
210	EST00168	20	TGTCAACTTCCCTTTGGCCT	GAAGCTTGCTCATTCAGGAA
136	EST00113	20	AL2 - TCGGAGAAGTTGCAGTTTCTG	GTTAAAAGCTGTTAGACGGGGC
120	EST00103	22	CACTGACTGACTCCTCTTTA	GGAACCGTAACTCTCCATAG
313	EST00276	X	ATTGACCTTCAATGTAATAA	TTGGATTGGGCAAAATAG

SUBSTITUTE SHEET

-51-

<u>SEQ ID</u>	<u>EST#</u>	<u>Chr</u>	<u>PRIMER #1</u>	<u>PRIMER #2</u>
162	EST00133	X	ATGTGAGCATCTATACCTGC	AATGAAGGCATGAGAATAGG
1669	EST00827	X	CGGACAAC TAGGATAAAATGC	TACGCGTTTGAATGGCTTGA
1917	EST01029	X	GAATAGCATTATTAGCCAGT	GGACCTATTGGAGATCTACT
1708	EST00858	X	AL2-AAGGCGAGGATTATGTGC	TTCTACTGGGTACACTTCGACC

Abbreviation: AL2: Amino-Link-2 Fluorescent Tag, Chr.: Chromosome.

SUBSTITUTE SHEET

The foregoing techniques have been used to further localize 9 ESTs and their associated genes to precise locations onto chromosome 6 or chromosome X, as reflected in Table 4A (in Example 7 below), using sublocalization techniques that employ somatic cell hybrids. ESTs were used as hybridization probes and mapped to other chromosomes using techniques disclosed in Example 7. Somatic cell hybrids were prepared that contained defined subsets of chromosomes 6 and X. Methods for preparing and selecting somatic cell hybrids are known in the art. For a review of an exemplary procedure to generate somatic cell hybrids containing the short arm of human chromosome 6, see Zoghbi, et al., *Genomics* 9(4):713-720 (1991). For a general review of somatic cell hybridization see Ledbetter et al. (*supra*). The hybrids were processed to obtain DNA and analyzed by PCR and by fluorescence in situ hybridization. SEQ ID NOs 19, 22, 1, 224, 288 mapped to chromosome 6, while SEQ ID NOs 162, 1917, 1699 and 1899 mapped to chromosome X using somatic cell hybrids.

EXAMPLE 6

Mapping of All ESTs to Human Chromosomes

The procedure of Example 5 is repeated for all of the ESTs from Examples 1 and 2 not previously mapped to human chromosomes. Data are generated corresponding to the data in Table 3 for all of the unmapped ESTs. As previously mentioned, virtually all of the ESTs will map to a unique chromosomal location. The inability of any ESTs to localize to a unique location will be readily ascertainable during the mapping process.

Physical mapping of the type reported in Table 4 on all the EST clones reported here would provide human chromosome markers spaced on average every 1.2 megabases and would roughly double the number of expressed sequences that have been localized to chromosomes (McKusick, V. FASEB

-53-

J. 5: 12-20 (1991)). Mapped ESTs are also a new resource to identify candidates for the estimated 5000 single-locus disease-associated genes (Id.).

EXAMPLE 7

5 **Alternative Technique for Mapping to Chromosomes**
 Mapping of ESTs to chromosomes using fluorescence in situ
 hybridization

10 This technique was used to map an EST to a particular location on a given chromosome. Cell cultures, tissue, or whole blood were used to obtain chromosomes.

 0.5 ml. of whole blood was added to RPMI 1640 and incubated 96 hours in a 5%CO₂/37°C incubator. 0.05 ug/ml colcemide was added to the culture one hour before harvest. Cells were collected and washed in PBS. The suspension was
15 incubated with a hypotonic solution of KCl added dropwise to reach a final volume of 5 ml. The cells were spun down and fixed by resuspending the cells in methanol and glacial acetic acid (3:1). The cell suspension was dropped onto glass slides and dried.

20 The slides were treated with RNase A and washed then dehydrated in a series of increasing concentrations of ethanol.

 The EST to be localized was nick-translated using fluorescently labeled nucleotide (Korenberg, Jr., et al.,
25 Cell 53(3):391-400 (1988)). Following nick translation, unincorporated label was removed by spin dialysis through Sepharose. The probe was further extracted with phenol-chloroform to remove additional protein. The chromosomes were denatured in formamide using techniques known in the art and the denatured probe was added to the slides. Following
30 hybridization, the cells were washed. The slides were studied under a fluorescent microscope. In addition, the chromosomes can be stained for G-banding or Q-banding using techniques known in the art.

-54-

The resulting metaphase chromosomes had fluorescent tags localized to those regions of the chromosome that were homologous to the EST. Thus, a particular EST was localized to a particular region on a given chromosome. In this manner, SEQ ID NOs 396, 485, 506, 1880 and 1894 were mapped using fluorescent in situ hybridization to locations on chromosomes 17, 7, 10 and 1 respectively (See Table 4B below). For a review of the technique see Verma et al., *Human Chromosomes: A Manual of Basic Techniques*. Pergamon Press, NY (1988), which is hereby incorporated by reference.

Table 4: Precise Chromosomal Localization of ESTs

	SEQ ID	EST#	Map Location
	-----	-----	-----
15	A.	19	EST00023 6p
		22	EST00301 6p
		1894	EST01643 6p21
		1	EST00007 6q
		224	EST00356 6q
		288	EST00219 6q
20		162	EST00133 Xp11.21 - Xp21.2
		1917	EST01029 Xp11.21 - Xp21.2
		1669	EST00827 Xq26 - Xq27.1
		1899	EST01014 Xq28
25	B.	1880	EST01634 1q32
		485	EST01466 7p13
		506	EST01471 10q11.2
		396	EST01443 17q25

EXAMPLE 8**Automated DNA Sequencing Accuracy**

ESTs that match human sequences in GenBank are excellent tools for the analysis of the accuracy of double-strand automated DNA sequencing. Ninety EST/GenBank matches were examined for the number of nucleotide mismatches and gaps required to achieve optimal alignment by the Genetics Computer Group (GCG) program BESTFIT (Devereux et al, *Nucleic Acids Research* 12: 387 (1984)).

-55-

The number of mismatches, insertions and deletions was counted for each hundred bases of the sequence (Table 5). As expected, the sequence quality was best closest to the primer and decreased rapidly after about 400 bases. The number of deletions and insertions relative to the GenBank reference sequence increased five- to ten-fold beyond 400 bases, while the number of mismatches doubled. The average accuracy rate for individual double-stranded sequencing runs was 97.7% to 400 bases.

TABLE 5. Accuracy Of Single-Run Double-Stranded Automated Sequencing

<u>Bases from Primer</u>	<u>Mismatches/ Ambiguities</u> ⁺	<u>Gaps Insertions</u> ⁺	<u>Percent Deletions</u> ⁺	<u>Aligned Accurate</u>	<u>Bases</u>
101 - 200	1.45	0.18	0.19	98.2	8,800
201 - 300	1.72	0.25	0.11	97.9	8,130
301 - 400	2.07	0.98	0.37	96.6	5,404
>400	3.53	2.63	1.06	92.8	3,197

ESTs statistically identical to known human sequences and those matching mitochondrial and ribosomal genes were aligned with sequenced from GenBank using the GCG program BESTFIT. The first 85 nucleotides was polylinker sequence which was not aligned with the pBluescript SK reference sequence. Tabulation of errors began 15 bases into the BESTFIT alignment and thus is reported beginning with bases 101-200. ⁺Error rates are reported as number of mismatches, insertions, or deletions per hundred aligned bases. "Mismatches" includes ambiguous base calls.

EXAMPLE 9

Probability of ESTs Containing Coding Sequences

The ESTs of the present invention were statistically evaluated using the coding-region prediction program CRM via the GRAIL server (Uberbacher, E. & Mural, R. Proc. Natl. Acad. Sci. USA, 88: 11261-5 (1991)). The CRM program uses a neural network to combine results from several different coding regions by looking at different 6 bp sequences found in coding exons and in introns. The program additionally conducts reading frame searches and assesses randomness at the third position of codons. This protocol categorizes sequences as having an excellent, good, marginal, or poor probability of containing coding regions. The results are reported in Tables 6-9. There were 219 ESTs categorized as "excellent" (Table 6); 120 categorized as "good" (Table 7); 113 categorized as "marginal" (Table 8); and 1743 categorized as "poor" (Table 9). These results indicate that most ESTs of the present invention comprise noncoding regions.

Table 6: ESTs with Excellent Probability of Containing Coding Sequence

SEQ ID#	EST#	SEQ ID#	EST#	SEQ ID#	EST#	SEQ ID#	EST#
7	EST00014	973	EST01987	1807	EST00941	2373	EST01393
15	EST00020	979	EST01993	1809	EST00943	2374	EST01394
48	EST00291	980	EST01994	1820	EST00951	2393	EST01417
62	EST00064	986	EST02000	1829	EST00958	2394	EST01418
66	EST00067	1000	EST02014	1849	EST00975	2396	EST01420
75	EST00074	1004	EST02018	1860	EST00983		
98	EST00260	1007	EST02021	1866	EST00989		
106	EST00092	1018	EST02032	1871	EST00994		
108	EST00094	1021	EST02035	1888	EST01005		
114	EST00098	1034	EST02050	1890	EST01007		
115	EST00099	1047	EST02063	1892	EST01009		
124	EST00107	1090	EST02109	1903	EST01018		
128	EST00252	1096	EST02115	1904	EST01019		
156	EST00130	1115	EST02135	1914	EST01026		
164	EST00135	1118	EST02138	1930	EST01040		
166	EST00137	1129	EST02149	1944	EST01050		
174	EST00296	1133	EST02153	1949	EST01054		
179	EST00145	1141	EST02163	1962	EST01062		
183	EST00148	1163	EST02187	1973	EST01071		
201	EST00163	1183	EST02208	1977	EST01075		
205	EST00165	1243	EST02272	1982	EST01080		
215	EST00172	1264	EST02293	1991	EST01088		
230	EST00181	1265	EST02294	1993	EST01090		
253	EST00199	1266	EST02295	2000	EST01097		
263	EST00203	1287	EST02317	2001	EST01098		
268	EST00369	1308	EST02338	2012	EST01106		
270	EST00207	1324	EST02354	2013	EST01107		
271	EST00283	1344	EST02374	2024	EST01117		
273	EST00208	1356	EST02386	2043	EST01131		
276	EST00211	1365	EST02396	2051	EST01138		
281	EST00214	1383	EST02415	2056	EST01142		
285	EST00286	1399	EST02433	2058	EST01144		
333	EST00394	1401	EST02435	2059	EST01145		
336	EST00397	1405	EST02439	2064	EST01149		
339	EST00400	1417	EST02452	2090	EST01167		
362	EST00418	1451	EST02487	2094	EST01171		
389	EST00440	1457	EST02493	2116	EST01192		
441	EST00481	1463	EST02500	2117	EST01193		
454	EST00493	1473	EST02510	2128	EST01202		
476	EST00509	1479	EST02516	2131	EST01205		
493	EST00522	1516	EST02555	2134	EST01208		
504	EST00529	1528	EST02569	2144	EST01216		
516	EST00538	1531	EST02572	2145	EST01217		
518	EST00540	1544	EST02586	2150	EST01222		
551	EST01482	1551	EST02593	2155	EST01227		
552	EST00565	1558	EST02601	2161	EST01231		
559	EST00570	1561	EST02604	2163	EST01238		
582	EST00592	1581	EST02625	2174	EST01242		
602	EST00606	1586	EST02631	2176	EST01244		
606	EST00609	1591	EST02636	2189	EST01255		
621	EST00620	1616	EST02661	2214	EST01272		
635	EST00629	1624	EST02670	2225	EST01278		
642	EST00634	1630	EST02676	2227	EST01279		
644	EST00636	1637	EST00796	2233	EST01284		
687	EST00671	1639	EST00799	2235	EST01286		
700	EST00683	1649	EST00808	2236	EST01287		
743	EST00714	1651	EST00810	2255	EST01302		
753	EST00721	1677	EST00835	2259	EST01304		
760	EST00726	1682	EST00839	2263	EST01307		
764	EST00729	1694	EST00849				
808	EST00761	1706	EST00857				
823	EST01864	1708	EST00858				
834	EST00771	1710	EST00860				
886	EST01886	1716	EST00865				
919	EST01921						
930	EST01933						
936	EST01939						
948	EST01957						
965	EST01978						

-59-

Table 7: ESTs with Good Probability of Containing Coding Sequence

<u>SEQ ID#</u>	<u>EST#</u>				
		1041	EST02057	2362	EST01383
		1083	EST02102	2378	EST01397
20	EST00024	1099	EST02118	2399	EST01423
72	EST00071	1105	EST02124	2407	EST02714
82	EST00078	1113	EST02133		
88	EST00084	1139	EST02161		
137	EST00272	1146	EST02168		
177	EST00328	1196	EST02221		
193	EST00156	1210	EST02238		
200	EST00162	1233	EST02262		
218	EST00175	1285	EST02314		
228	EST00179	1331	EST02361		
247	EST00279	1388	EST02421		
264	EST00204	1418	EST02453		
267	EST00297	1439	EST02475		
296	EST00228	1502	EST02540		
371	EST00426	1537	EST02578		
385	EST00436	1563	EST02606		
392	EST00442	1599	EST02644		
414	EST00460	1602	EST02647		
433	EST00474	1693	EST00848		
453	EST00492	1695	EST00850		
471	EST00505	1729	EST00877		
496	EST00525	1730	EST00878		
524	EST00544	1738	EST00883		
526	EST00546	1739	EST00885		
529	EST00549	1743	EST00888		
549	EST00563	1768	EST00908		
557	EST00569	1780	EST00916		
578	EST00588	1804	EST00938		
596	EST00602	1805	EST00939		
607	EST00610	1811	EST00945		
619	EST00619	1819	EST00950		
657	EST00646	1826	EST00956		
660	EST00649	1830	EST00959		
689	EST00673	1845	EST00971		
695	EST00679	1848	EST00974		
699	EST00682	1853	EST00977		
729	EST00703	1967	EST01066		
742	EST00713	1992	EST01089		
747	EST00717	1994	EST01091		
755	EST00723	<u>SEQ ID#</u>	<u>EST#</u>		
759	EST00725				
776	EST00738	1997	EST01094		
778	EST00740	2046	EST01134		
782	EST01551	2101	EST01177		
829	EST00768	2102	EST01178		
835	EST00772	2105	EST01181		
836	EST00773	2106	EST01182		
862	EST01872	2141	EST01213		
881	EST01881	2184	EST01251		
<u>SEQ ID#</u>	<u>EST#</u>	2196	EST01260		
		2203	EST01264		
884	EST01884	2232	EST01283		
924	EST01926	2308	EST01339		
929	EST01932	2345	EST01368		
938	EST01941	2346	EST01369		
971	EST01985	2351	EST01373		
995	EST02009	2354	EST01375		
996	EST02010	2355	EST01376		
1031	EST02046	2359	EST01380		

Table 8: ESTs with Marginal Probability of Containing Coding Sequence

<u>SEQ ID#</u>	<u>EST#</u>		
11	EST00018	1222	EST02251
12	EST00274	1224	EST02253
24	EST00027	1228	EST02257
45	EST00364	1267	EST02296
79	EST00076	1301	EST02331
90	EST00302	1397	EST02431
110	EST00096	1448	EST02484
144	EST00120	1480	EST02517
145	EST00121	1493	EST02531
192	EST00155	1499	EST02537
222	EST00177	1503	EST02541
234	EST00184	1527	EST02568
277	EST00212	1536	EST02577
319	EST00381	1548	EST02590
368	EST00423	1562	EST02605
370	EST00425	1572	EST02615
387	EST00438	1575	EST02618
402	EST00451	1595	EST02640
415	EST00461	1608	EST02653
418	EST00464	1610	EST02655
426	EST00470	1621	EST02667
503	EST00528	1627	EST02674
517	EST00539	1629	EST02677
522	EST00543	1631	EST02678
532	EST00551	1683	EST00840
540	EST00557	1692	EST00847
570	EST00580	1751	EST00895
573	EST00583	1756	EST00900
576	EST00586	1764	EST02690
613	EST00615	1770	EST00910
617	EST00617	1793	EST00929
626	EST00622	1847	EST00973
681	EST00665	1877	EST00998
726	EST00700	1897	EST01012
727	EST00701	1900	EST01015
738	EST00711	1939	EST01655
745	EST00715	1940	EST01046
752	EST00720	1954	EST01058
791	EST00746	<u>SEQ ID#</u>	<u>EST#</u>
795	EST00749	1990	EST01087
803	EST00756	2008	EST01103
845	EST00777	2031	EST01123
852	EST00782	2041	EST01130
854	EST00784	2044	EST01132
907	EST01907	2060	EST01146
912	EST01912	2100	EST01176
935	EST01938	2136	EST01210
<u>SEQ ID#</u>	<u>EST#</u>	2153	EST01225
968	EST01981	2204	EST01265
985	EST01999	2212	EST01270
988	EST02002	2248	EST01297
1043	EST02059	2250	EST01299
1081	EST02100	2266	EST01310
1089	EST02108	2309	EST01340
1116	EST02136	2347	EST01370
1134	EST02154	2388	EST01406
1205	EST02233	2398	EST01422
		2405	EST01427

SEQ ID#	EST#	103	EST00317	204	EST00235	309	EST00174	404	EST00453
1	EST00007	104	EST00354	206	EST00166	315	EST00008	405	EST00454
2	EST00009	105	EST00365	207	EST00167	316	EST00378	406	EST00455
3	EST00010	107	EST00093	209	EST00331	317	EST00379	407	EST00456
4	EST00011	109	EST00095	210	EST00168	318	EST00380	408	EST00457
5	EST00012	111	EST00281	211	EST00332	320	EST00382	409	EST01444
6	EST00013	112	EST00318	212	EST00169	321	EST00383	410	EST00458
8	EST00234	113	EST00097	213	EST00170	322	EST00384	411	EST00459
10	EST00016	116	EST00100	214	EST00171	323	EST00385	412	EST01445
14	EST00019	117	EST00319	216	EST00173	325	EST00386	416	EST00462
16	EST00021	118	EST00101	219	EST00176	326	EST00387	417	EST00463
17	EST00022	119	EST00102	220	EST00372	327	EST00388	419	EST00465
18	EST00373	120	EST00103	221	EST00359	328	EST00389	420	EST00466
19	EST00023	121	EST00104	224	EST00356	329	EST00390	421	EST00467
21	EST00025	122	EST00105	225	EST00178	330	EST00391	422	EST01447
23	EST00026	123	EST00106	226	EST00333	331	EST00392	423	EST00468
25	EST00028	125	EST00108	229	EST00180	332	EST00393	424	EST01448
27	EST00029	126	EST00109	231	EST00334	334	EST00395	425	EST00469
28	EST00030	127	EST00320	232	EST00182	335	EST00396	427	EST01449
29	EST00031	129	EST00321	233	EST00183	337	EST00398	428	EST01451
30	EST00032	130	EST00355	235	EST00185	340	EST00402	429	EST00471
31	EST00033	131	EST00322	236	EST00186	341	EST00403	431	EST00473
32	EST00233	133	EST00111	237	EST00187	342	EST00404	432	EST01452
33	EST00034	134	EST00375	238	EST00188	344	EST00405	434	EST00475
34	EST00035	135	EST00112	239	EST00189	345	EST00406	435	EST00476
35	EST00036	136	EST00113	240	EST00335	347	EST01829	436	EST00477
36	EST00037	138	EST00114	241	EST00191	348	EST01830	437	EST00478
39	EST00039	139	EST00116	242	EST00192	349	EST01831	438	EST00479
40	EST00040	140	EST00117	243	EST00193	350	EST00407	439	EST00480
41	EST00041	141	EST00118	244	EST00194	351	EST00408	440	EST01454
42	EST00042	142	EST00323	245	EST00347	352	EST00409	442	EST01456
46	EST00044	143	EST00119	246	EST00196	353	EST00410	443	EST00482
47	EST00046	146	EST00122	250	EST00197	354	EST01433	444	EST00483
49	EST00047	147	EST00292	252	EST00198	355	EST00411	446	EST00485
50	EST00048	148	EST00236	254	EST00200	356	EST00412	447	EST00486
51	EST00049	149	EST00123	255	EST00201	357	EST00413	448	EST00487
52	EST00052	150	EST00124	256	EST00345	358	EST00414	449	EST00488
53	EST00054	151	EST00125	257	EST00337	359	EST00415	450	EST00489
54	EST00055	152	EST00126	259	EST00202	360	EST00416	451	EST00490
55	EST00056	153	EST00127	260	EST00357	361	EST00417	452	EST00491
56	EST00057	154	EST00128	261	EST00338	363	EST00419	455	EST00494
57	EST00058	155	EST00129	262	EST00339	364	EST00420	457	EST00495
58	EST00059	157	EST00131	265	EST00205	365	EST01434	458	EST00496
59	EST00061	158	EST00132	266	EST00206	366	EST00421	459	EST00497
60	EST00062	159	EST00325	272	EST00340	367	EST00422	460	EST01457
63	EST00065	160	EST00326	274	EST00268	369	EST00424	461	EST01836
64	EST00066	162	EST00133	275	EST00209	372	EST00427	462	EST00498
67	EST00351	163	EST00134	278	EST00342	373	EST01832	464	EST00499
68	EST00068	165	EST00136	279	EST00213	374	EST00428	465	EST00500
69	EST00360	167	EST00138	280	EST00343	375	EST00429	466	EST00501
71	EST00070	168	EST00140	283	EST00215	376	EST01436	467	EST00502
73	EST00072	169	EST00141	284	EST00216	377	EST00430	468	EST00503
74	EST00073	170	EST00295	286	EST00217	378	EST00431	470	EST00504
76	EST00075	171	EST00327	287	EST00218	379	EST00432	SEQ ID#	EST#
80	EST00077	172	EST00142	288	EST00219	380	EST01439	473	EST00506
81	EST00315	173	EST00143	289	EST00220	381	EST00433	474	EST00507
83	EST00079	175	EST00144	290	EST00221	382	EST00434	477	EST01463
84	EST00080	178	EST00294	291	EST00222	SEQ ID#	EST#	478	EST00510
85	EST00081	182	EST00329	292	EST00223	383	EST00435	479	EST00511
86	EST00082	184	EST00149	293	EST00224	384	EST01440	480	EST01464
87	EST00083	185	EST00150	294	EST00225	386	EST00437	481	EST00512
89	EST00085	186	EST00151	SEQ ID#	EST#	388	EST00439	482	EST01465
91	EST00086	190	EST00153	295	EST00226	390	EST01442	483	EST00513
92	EST00087	191	EST00154	297	EST00230	391	EST00441	484	EST00514
94	EST00353	194	EST00157	298	EST00231	393	EST00443	487	EST00516
95	EST00088	SEQ ID#	EST#	302	EST00303	395	EST00445	488	EST00517
96	EST00089	195	EST00158	303	EST00348	397	EST00446	489	EST00518
99	EST00316	196	EST00159	304	EST00307	398	EST00447	490	EST00519
SEQ ID#	EST#	197	EST00160	305	EST00308	399	EST00448	491	EST00520
100	EST00090	198	EST00161	306	EST00309	400	EST00449	492	EST00521
101	EST00091	199	EST00277	307	EST00312	401	EST00450	495	EST00524
		203	EST00164	308	EST00314	403	EST00452	497	EST00526

498	EST01467	600	EST01492	697	EST00680	799	EST00752	894	EST01894
499	EST01468	601	EST01493	698	EST00681	800	EST00753	895	EST01895
500	EST00527	603	EST01494	701	EST01522	801	EST00754	896	EST01896
501	EST02715	604	EST00607	702	EST00684	804	EST00757	897	EST01897
502	EST01469	605	EST00608	703	EST00685	805	EST00758	898	EST01898
507	EST00530	609	EST01496	704	EST00686	806	EST00759	899	EST01899
508	EST00531	610	EST00612	705	EST00687	807	EST00760	900	EST01900
509	EST01472	611	EST00613	706	EST00688	809	EST00762	901	EST01901
510	EST00532	612	EST00614	708	EST00689	810	EST00763	902	EST01902
511	EST00533	615	EST00616	709	EST00690	811	EST00764	903	EST01903
512	EST00534	616	EST01497	710	EST00691	813	EST00765	904	EST01904
513	EST00535	618	EST01498	711	EST00692	814	EST00766	905	EST01905
514	EST00536	620	EST01499	712	EST00693	815	EST01855	906	EST01906
515	EST00537	622	EST01843	713	EST00694	816	EST01856	908	EST01908
519	EST00541	623	EST00621	714	EST00695	817	EST01857	909	EST01909
520	EST00542	624	EST01500	715	EST01523	818	EST01858	910	EST01910
521	EST01474	625	EST01844	716	EST01524	819	EST01859	911	EST01911
523	EST01838	627	EST00623	717	EST01525	820	EST01860	914	EST01914
525	EST00545	628	EST01503	718	EST00696	822	EST01863	915	EST01915
527	EST00547	629	EST00624	719	EST01526	825	EST01866	916	EST01917
528	EST00548	630	EST01505	720	EST00697	826	EST01867	917	EST01919
530	EST01477	631	EST00625	721	EST01527	827	EST01558	918	EST01920
531	EST00550	632	EST00626	722	EST01528	828	EST00767	920	EST01922
533	EST00552	633	EST00627	723	EST00698	830	EST01559	921	EST01923
534	EST01478	634	EST00628	725	EST00699	831	EST00769	922	EST01924
535	EST00553	636	EST01507	728	EST00702	832	EST00770	923	EST01925
536	EST01479	637	EST00630	730	EST00704	837	EST01561	925	EST01927
537	EST00554	638	EST00631	731	EST00705	838	EST00774	926	EST01929
538	EST00555	640	EST01509	732	EST00706	839	EST01562	927	EST01930
539	EST00556	641	EST00633	733	EST00707	840	EST00775	928	EST01931
541	EST00558	643	EST00635	734	EST00708	841	EST00776	931	EST01934
542	EST01480	645	EST00637	735	EST00709	842	EST01563	932	EST01935
543	EST00559	646	EST00638	736	EST01532	843	EST01564	933	EST01936
544	EST00560	647	EST00639	737	EST00710	844	EST01565	934	EST01937
545	EST01481	648	EST00640	739	EST01534	846	EST00778	937	EST01940
547	EST00561	649	EST00641	740	EST01535	847	EST00779	939	EST01943
548	EST00562	651	EST00643	741	EST00712	848	EST01566	SEQ ID#	EST#
550	EST00564	652	EST01510	744	EST01537	849	EST01567	940	EST01944
553	EST00566	654	EST00644	746	EST00716	850	EST00780	941	EST01945
555	EST01483	655	EST00645	748	EST01850	851	EST00781	942	EST01947
556	EST00568	656	EST01513	749	EST00719	SEQ ID#	EST#	943	EST01948
558	EST01484	658	EST00647	750	EST01539	853	EST00783	944	EST01949
560	EST01485	659	EST00648	751	EST01540	855	EST00785	945	EST01950
561	EST00571	661	EST00650	754	EST00722	856	EST01568	946	EST01953
562	EST00572	662	EST00651	SEQ ID#	EST#	857	EST01868	947	EST01954
563	EST00573	663	EST00652	756	EST01541	858	EST01869	949	EST01958
564	EST00574	664	EST00653	758	EST00724	859	EST01870	950	EST01959
565	EST00575	665	EST00654	761	EST01544	860	EST00786	953	EST01962
566	EST00576	SEQ ID#	EST#	762	EST00727	861	EST01871	954	EST01963
567	EST00577	666	EST01514	763	EST00728	863	EST01873	956	EST01968
568	EST00578	667	EST00655	765	EST00730	864	EST00787	957	EST01969
569	EST00579	668	EST00656	766	EST00731	865	EST01569	958	EST01970
SEQ ID#	EST#	669	EST00657	767	EST00732	866	EST01874	959	EST01972
571	EST00581	670	EST00658	768	EST00733	867	EST01875	960	EST01973
572	EST00582	671	EST00659	770	EST00735	868	EST01876	961	EST01974
574	EST00584	672	EST00660	771	EST01546	869	EST00788	962	EST01975
575	EST00585	673	EST01515	772	EST00736	870	EST00789	963	EST01976
577	EST00587	674	EST01516	774	EST01548	871	EST00790	964	EST01977
580	EST00590	675	EST00661	775	EST00737	872	EST00791	966	EST01979
581	EST00591	676	EST00662	777	EST00739	873	EST00792	967	EST01980
583	EST00593	677	EST00663	779	EST00741	874	EST00793	970	EST01983
584	EST00594	678	EST01517	780	EST01549	875	EST00794	972	EST01986
585	EST00595	679	EST01518	781	EST01550	876	EST00795	974	EST01988
586	EST00596	680	EST00664	783	EST01552	877	EST01877	975	EST01989
587	EST01488	682	EST00666	785	EST01553	878	EST01878	976	EST01990
588	EST00597	683	EST00667	786	EST00742	879	EST01879	977	EST01991
589	EST00598	684	EST00668	787	EST00743	880	EST01880	978	EST01992
590	EST00599	685	EST00669	788	EST00744	882	EST01882	981	EST01995
591	EST01489	686	EST00670	789	EST00745	883	EST01883	982	EST01996
592	EST00600	688	EST00672	790	EST01554	885	EST01885	983	EST01997
593	EST00601	690	EST00674	792	EST00747	887	EST01887	984	EST01998
595	EST01840	692	EST00676	793	EST00748	889	EST01889	987	EST02001
597	EST00603	693	EST00677	794	EST01555	890	EST01890	989	EST02003
598	EST00604	694	EST00678	796	EST00750	892	EST01892	990	EST02004
599	EST00605	696	EST01521	797	EST00751	893	EST01893	991	EST02005

-63-

992	EST02006	1086	EST02105	1184	EST02209	1274	EST02303	1363	EST02394
994	EST02008	1087	EST02106	1185	EST02210	1275	EST02304	1364	EST02395
997	EST02011	1088	EST02107	1186	EST02211	1276	EST02305	1366	EST02397
999	EST02013	1091	EST02110	1187	EST02212	1278	EST02307	1367	EST02398
1001	EST02015	1093	EST02112	1188	EST02213	1279	EST02308	1368	EST02399
1002	EST02016	1095	EST02114	1189	EST02214	1280	EST02309	1370	EST02401
1003	EST02017	1097	EST02116	1190	EST02215	1281	EST02310	1372	EST02403
1005	EST02019	1098	EST02117	1191	EST02216	1282	EST02311	1373	EST02404
1006	EST02020	1100	EST02119	1192	EST02217	1283	EST02312	1375	EST02406
1008	EST02022	1101	EST02120	1193	EST02218	1284	EST02313	1376	EST02407
1009	EST02023	1102	EST02121	1194	EST02219	1286	EST02316	1377	EST02408
1010	EST02024	1104	EST02123	1195	EST02220	1288	EST02318	1378	EST02409
1011	EST02025	1106	EST02125	1197	EST02222	1289	EST02319	1379	EST02410
1012	EST02026	1107	EST02126	1198	EST02223	1290	EST02320	1380	EST02411
1013	EST02027	1108	EST02127	1199	EST02224	1291	EST02321	1381	EST02413
1014	EST02028	1109	EST02128	1200	EST02226	1292	EST02322	1382	EST02414
1015	EST02029	1110	EST02129	1201	EST02228	1293	EST02323		
1016	EST02030	1111	EST02131	1202	EST02229	1294	EST02324		
1017	EST02031	1112	EST02132	1203	EST02230	1295	EST02325		
1019	EST02033	1114	EST02134	1204	EST02232	1296	EST02326		
1022	EST02036	1117	EST02137	1206	EST02234	SEQ ID#	EST#		
1023	EST02037	1119	EST02139	1207	EST02235				
1024	EST02038	1120	EST02140	1208	EST02236	1298	EST02328		
1025	EST02040	1121	EST02141	1209	EST02237	1299	EST02329		
1026	EST02041	1122	EST02142	SEQ ID#	EST#	1300	EST02330		
1027	EST02042	1123	EST02143			1302	EST02332		
1028	EST02043	1124	EST02144	1211	EST02239	1303	EST02333		
1029	EST02044	1125	EST02145	1212	EST02240	1304	EST02334		
1030	EST02045	SEQ ID#	EST#	1213	EST02241	1305	EST02335		
1032	EST02048			1214	EST02242	1306	EST02336		
1033	EST02049	1127	EST02147	1215	EST02244	1307	EST02337		
1036	EST02052	1128	EST02148	1216	EST02245	1309	EST02339		
SEQ ID#	EST#	1130	EST02150	1217	EST02246	1310	EST02340		
		1131	EST02151	1218	EST02247	1311	EST02341		
1037	EST02053	1132	EST02152	1219	EST02248	1313	EST02343		
1038	EST02054	1135	EST02155	1220	EST02249	1314	EST02344		
1040	EST02056	1136	EST02156	1221	EST02250	1315	EST02345		
1042	EST02058	1137	EST02157	1223	EST02252	1316	EST02346		
1044	EST02060	1138	EST02159	1225	EST02254	1317	EST02347		
1045	EST02061	1140	EST02162	1226	EST02255	1318	EST02348		
1046	EST02062	1142	EST02164	1227	EST02256	1319	EST02349		
1048	EST02064	1143	EST02165	1232	EST02261	1320	EST02350		
1049	EST02065	1144	EST02166	1234	EST02263	1321	EST02351		
1050	EST02066	1145	EST02167	1235	EST02264	1322	EST02352		
1051	EST02067	1148	EST02170	1236	EST02265	1323	EST02353		
1052	EST02068	1149	EST02171	1237	EST02266	1325	EST02355		
1053	EST02069	1150	EST02172	1238	EST02267	1326	EST02356		
1054	EST02070	1152	EST02174	1239	EST02268	1327	EST02357		
1055	EST02071	1153	EST02175	1240	EST02269	1328	EST02358		
1056	EST02072	1154	EST02176	1241	EST02270	1329	EST02359		
1057	EST02073	1155	EST02177	1242	EST02271	1330	EST02360		
1058	EST02074	1156	EST02178	1244	EST02273	1333	EST02363		
1059	EST02075	1157	EST02180	1246	EST02275	1334	EST02364		
1060	EST02076	1158	EST02181	1247	EST02276	1335	EST02365		
1061	EST02078	1159	EST02182	1248	EST02277	1336	EST02366		
1062	EST02079	1160	EST02183	1249	EST02278	1337	EST02367		
1063	EST02081	1161	EST02184	1250	EST02279	1338	EST02368		
1064	EST02082	1162	EST02185	1251	EST02280	1339	EST02369		
1065	EST02083	1164	EST02188	1252	EST02281	1342	EST02372		
1066	EST02084	1165	EST02189	1253	EST02282	1343	EST02373		
1067	EST02085	1166	EST02190	1254	EST02283	1345	EST02375		
1068	EST02086	1167	EST02191	1255	EST02284	1346	EST02376		
1070	EST02088	1168	EST02193	1256	EST02285	1347	EST02377		
1071	EST02089	1169	EST02194	1257	EST02286	1349	EST02379		
1072	EST02090	1170	EST02195	1258	EST02287	1350	EST02380		
1073	EST02091	1171	EST02196	1259	EST02288	1351	EST02381		
1074	EST02092	1172	EST02197	1260	EST02289	1352	EST02382		
1075	EST02093	1173	EST02198	1261	EST02290	1353	EST02383		
1076	EST02094	1174	EST02199	1262	EST02291	1354	EST02384		
1077	EST02096	1175	EST02200	1263	EST02292	1355	EST02385		
1078	EST02097	1176	EST02201	1268	EST02297	1357	EST02387		
1079	EST02098	1177	EST02202	1269	EST02298	1358	EST02388		
1080	EST02099	1178	EST02203	1270	EST02299	1359	EST02390		
1082	EST02101	1179	EST02204	1271	EST02300	1360	EST02391		
1084	EST02103	1180	EST02205	1272	EST02301	1361	EST02392		
1085	EST02104	1182	EST02207	1273	EST02302	1362	EST02393		

1907	EST01022	2016	EST01110	2118	EST01194	2223	EST01742	2332	EST01794
1908	EST01023	2018	EST01111	2119	EST01195	2224	EST01277	2333	EST01357
1909	EST01024	2019	EST01112	2122	EST01197	2228	EST01280	2335	EST01359
1911	EST02694	2020	EST01113	2123	EST01713	2229	EST01281	2336	EST01360
1912	EST01025	2021	EST01114	2124	EST01198	2231	EST01746	2337	EST01361
1913	EST01646	2022	EST01115	2125	EST01199	2237	EST01288	2340	EST01802
1915	EST01027	2023	EST01116	2126	EST01200	2238	EST01289	2341	EST01364
1916	EST01028	2025	EST01118	2127	EST01201	2239	EST01290	2343	EST01366
1917	EST01029	2026	EST01119	2129	EST01203	2240	EST01291	2344	EST01367
1918	EST02695	2027	EST01120	2130	EST01204	2241	EST01747	2349	EST01372
1919	EST01030	2028	EST01121	2132	EST01206	2242	EST01292	2350	EST02708
1920	EST01031	2029	EST01682	2133	EST01207	2243	EST01293	2352	EST01374
1921	EST01647	2030	EST01122	2135	EST01209	2244	EST01294	2356	EST01377
1922	EST01032	2033	EST01684	2137	EST01211	2246	EST01295	2357	EST01378
1923	EST01033	2034	EST01124	2139	EST01716	2247	EST01296	2360	EST01381
1924	EST01034	2035	EST01125	2140	EST01212	2249	EST01298	2361	EST01382
1925	EST01035	2036	EST01126	2142	EST01214	2251	EST01300	2363	EST01384
1926	EST01036	2037	EST01686	2143	EST01215	2252	EST01750	2364	EST01385
1927	EST01037	2038	EST01127	2147	EST01219	2253	EST01301	2365	EST01386
1929	EST01039	2039	EST01128	2148	EST01220	2256	EST02718	2366	EST01387
1932	EST01042	2040	EST01129	2151	EST01223	2257	EST01303	2369	EST01811
1934	EST01043	2042	EST01688	2152	EST01224	2258	EST01754	2370	EST01390
1935	EST01044	2045	EST01133	2154	EST01226	2260	EST01305	2371	EST01391
1936	EST01045	2047	EST01135	2156	EST01718	2261	EST01755	2372	EST01392
1937	EST01652	2048	EST01136	2157	EST01719	2262	EST01306	2375	EST01815
1938	EST01654	2049	EST01689	2158	EST01228	2264	EST01308	2376	EST01395
1941	EST01047	2050	EST01137	2159	EST01229	2265	EST01309	2377	EST01396
1942	EST01048	2052	EST01139	2160	EST01230	2268	EST01311	2379	EST01398
1943	EST01049	2053	EST01140	2162	EST01232	2269	EST01312	2380	EST01399
1945	EST01051	2054	EST01141	2163	EST01233	2270	EST01313	2381	EST01400
1946	EST02696	2055	EST01690	2164	EST01234	2271	EST01314	2382	EST01401
1947	EST01052	2057	EST01143	2165	EST01720	2272	EST01762	2383	EST01402
1948	EST01053	2061	EST01147	2166	EST01236	2273	EST01315	2384	EST01403
1950	EST01055	2062	EST02701	2167	EST01237	2275	EST01316	2385	EST01816
1951	EST01056	2063	EST01148	2169	EST01722	2276	EST01317	2386	EST01404
1952	EST01057	2065	EST01691	2170	EST01239	2277	EST01318	2387	EST01405
1955	EST01662	2066	EST01692	2171	EST01240	2278	EST01319		
1957	EST01059	2067	EST01693	2172	EST01241	2279	EST01320		
1958	EST01060	2069	EST01150	2175	EST01243	2280	EST01763		
1959	EST01061	2070	EST01151	2177	EST01245	2284	EST01323		
1963	EST01063	2072	EST01152	2178	EST01726	SEQ ID#	EST#		
1964	EST01064	2074	EST01698	2179	EST01246				
1966	EST01065	2075	EST01153	2180	EST01247	2285	EST01768		
1968	EST01067	2076	EST02702	2181	EST01248	2287	EST01770		
1969	EST01068	2077	EST01154	SEQ ID#	EST#	2288	EST01324		
1970	EST01666	2078	EST01155			2290	EST01772		
1971	EST01069	2079	EST01156	2182	EST01249	2291	EST01773		
1972	EST01070	2080	EST01157	2183	EST01250	2292	EST01326		
1975	EST01073	SEQ ID#	EST#	2185	EST01252	2293	EST01327		
1976	EST01074			2186	EST01253	2294	EST01328		
1978	EST01076	2081	EST01158	2187	EST01727	2295	EST01329		
1979	EST01077	2082	EST01159	2188	EST01254	2296	EST01330		
SEQ ID#	EST#	2083	EST01160	2190	EST01728	2298	EST01331		
1980	EST01078	2084	EST01161	2191	EST01256	2299	EST01332		
1981	EST01079	2085	EST01162	2193	EST01258	2301	EST01334		
1983	EST01081	2086	EST01163	2194	EST01729	2304	EST01780		
1984	EST01082	2087	EST01164	2195	EST01259	2305	EST01336		
1985	EST01083	2088	EST01166	2197	EST01261	2306	EST01337		
1986	EST01084	2091	EST01168	2198	EST01730	2310	EST01341		
1988	EST01085	2093	EST01170	2199	EST01262	2311	EST01342		
1989	EST01086	2095	EST01701	2200	EST01731	2312	EST01343		
1995	EST01092	2096	EST01172	2201	EST01263	2313	EST01344		
1996	EST01093	2097	EST01173	2202	EST01732	2315	EST01346		
1998	EST01095	2098	EST01174	2205	EST01735	2316	EST01782		
1999	EST01096	2099	EST01175	2206	EST01736	2317	EST01347		
2002	EST01099	2103	EST01179	2208	EST01267	2318	EST01348		
2003	EST01675	2104	EST01180	2209	EST02717	2319	EST01349		
2005	EST01100	2107	EST01183	2210	EST01268	2321	EST01350		
2006	EST01101	2108	EST01184	2211	EST01269	2322	EST01351		
2007	EST01102	2109	EST01185	2213	EST01271	2323	EST01789		
2009	EST01677	2110	EST01186	2215	EST01273	2325	EST01353		
2010	EST01104	2111	EST01187	2218	EST01274	2327	EST01354		
2011	EST01105	2112	EST01188	2219	EST01275	2328	EST01355		
2014	EST01108	2113	EST01189	2220	EST01740	2329	EST01792		
2015	EST01109	2114	EST01190	2221	EST01741	2330	EST01793		
		2115	EST01191	2222	EST01276	2331	EST01356		

-66-

<u>SEQ ID#</u>	<u>EST#</u>
2389	EST01407
2391	EST01415
2392	EST01416
2395	EST01419
2397	EST01421
2401	EST01424
2403	EST01425
2404	EST01426
2406	EST02713
2409	EST00273

SUBSTITUTE SHEET

-67-

EXAMPLE 10

Functional Groupings of ESTs and Corresponding Genes

By matching new human ESTs to known sequences from other species, the apparent function of the gene corresponding to the EST can be ascertained. The data generated in Example 3 and 4 have been used to categorize 127 of the ESTs of the present invention, and their corresponding genes, into predicted functional groups. (These 127 are ESTs with database matches to sequences from other species for which a function was known.) Two different grouping schemes have been used.

The first scheme separates the sequences into three broad categories: metabolic; regulatory; and structural. These groupings are set out in Table 10.

The second grouping scheme separates the sequences into 13 specific categories: cell surface proteins; developmental control; energy metabolism; kinases and phosphatases; oncogenes; other metabolism-related polypeptides; peptidases and peptidase inhibitors; receptors; structural and cytoskeletal; signal transduction; transporters; transcription, translation, and subcellular localization; and transcription factors. These groupings are set out in Table 11.

Table 10: Three-Class Functional Groupings of ESTs

SEQ ID	EST#	Group	Putative Identification
1834	EST01620	M	AMP deaminase, brain
97	EST00289	M	Aconitase
691	EST00675	M	Alcohol dehydrogenase
2092	EST01700	M	Anion exchanger homolog AE3
396	EST01443	M	CDPdiacylglycerol-serine O-phosphatidyltransferase
1956	EST01663	M	Ca ²⁺ -transporting ATPase 2
1039	EST02055	M	Calcium channel
2192	EST01257	M	Diacylglycerol kinase, lymphocyte
1441	EST02477	M	Diamine acetyltransferase
2289	EST01325	M	Fatty acid synthase
310	EST00377	M	Fo ATPase beta subunit, mitochondrial
1667	EST00825	M	Gamma-aminobutyric acid transporter
1412	EST02446	M	Glutamate-aspartate carrier protein
1020	EST02034	M	Glutaminase
2326	EST01791	M	Inositol-1,4,5-trisphosphate 3-kinase
2173	EST01724	M	Lon protease
1427	EST02463	M	Long-chain-fatty-acid-CoA ligase
2226	EST01744	M	NAD(P) ⁺ transhydrogenase (B-specific)
1566	EST02609	M	Neutrophil oxidase factor
1681	EST01573	M	Nucleoside diphosphate kinase
2254	EST01751	M	Phosphatidylinositol-4,5-bisphosphate phosphodiesterase
93	EST00287	M	Processing enhancing protein
2297	EST01775	M	Prohormone cleavage enzyme
9	EST00376	M	Prolyl endopeptidase
1654	EST01572	M	Protochlorophyllide reductase
38	EST00374	M	RNA polymerase II 6th subunit (RPO26)
1715	EST01583	M	Ribosomal protein L18a
1856	EST01627	M	Ribosomal protein L1a
1974	EST01667	M	Ribosomal protein L3
301	EST00300	M	Ribosomal protein L30
22	EST00301	M	Ribosomal protein S10
2402	EST01826	M	Ribosomal protein S10
463	EST01459	M	Ribosomal protein YL10
2073	EST01697	M	Succinate dehydrogenase flavoprotein
2138	EST01715	M	Succinate dehydrogenase flavoprotein
1771	EST01601	M	Thiosulfate sulfurtransferase (rhodanese)
2121	EST01711	M	Valine-tRNA ligase
1726	EST01588	M	XPR2 alkaline extracellular protease
913	EST01913	M	Clathrin coat assembly protein AP50 homolog
1035	EST02051	M	J1 protein
969	EST01982	R	ADP-ribosylation factor 1
1126	EST02146	R	Calbindin D28
1910	EST01645	R	Calmodulin
485	EST01466	R	Calmodulin-dependent protein kinase, type II, beta
2302	EST01779	R	Discs-large tumor suppressor
188	EST00256	R	Enhancer of split
1229	EST02258	R	KUP protein
993	EST02007	R	Kinase 5 protein
2282	EST01764	R	Lamin B receptor
SEQ ID	EST#	Group	Putative Identification
161	EST00247	R	MARCKS (myristoylated alanine-rich protein kinase)
769	EST00734	R	MARCKS homolog
1386	EST02418	R	MARCKS homolog
227	EST00259	R	Notch/Xotch
952	EST01961	R	Notch/Xotch
1395	EST02429	R	Nuclear factor 1-like protein (NF1)
2353	EST01806	R	Prohibitin
1069	EST02087	R	Protein kinase C, zeta
1933	EST01650	R	Protein phosphatase 2A beta subunit

-69-

202	EST00298	R	Protein-tyrosine phosphatase LRP
1478	EST02515	R	Rab5
1408	EST02442	R	Seven in absentia
300	EST00232	R	Transforming protein (dbl)
1147	EST02169	R	Tyrosine kinase
1348	EST02378	R	cAMP-dependent protein kinase inhibitor
1931	EST01041	R	cAMP-regulated phosphoprotein
1413	EST02447	R	cAMP-specific phosphodiesterase
37	EST00038	R	ras p21-like small GTP-binding protein (smg GDS)
102	EST00248	R	rho H12/ ARH12
299	EST00249	R	smg p25A GDP dissociation inhibitor
189	EST00282	R	trkB
1332	EST02362	R	GA binding protein, beta subunit
1277	EST02306	R	Bib protein
43	EST00371	R	Maternal G10 protein
1704	EST01580	R	Myeloid differentiation primary response gene My
346	EST01828	R	Otd homeotic protein
187	EST00152	R	Wilm's tumor-related protein
249	EST00275	R	Zinc Finger Proteins
413	EST01446	R	Zinc Finger Proteins
469	EST01460	R	Zinc Finger Proteins
833	EST01560	R	Zinc Finger Proteins
1230	EST02259	R	Zinc finger proteins
1496	EST02534	R	Zinc finger proteins
2324	EST01352	R	Zinc Finger Proteins
208	EST00250	S	60K filarial antigen
2320	EST01784	S	60K filarial antigen
251	EST00370	S	Actin, other
2146	EST01218	S	Actin, other
248	EST00271	S	Actinin, alpha
891	EST01891	S	Actinin, alpha
1500	EST02538	S	Actinin, alpha
132	EST00110	S	Agrin
1852	EST01625	S	Agrin
1965	EST01664	S	Amyloid A4
2068	EST01694	S	Amyloid A4
2408	EST00244	S	Amyloid A4
1880	EST01634	S	Axonal glycoprotein TAG-1
2004	EST01676	S	Cofilin
650	EST00642	S	Dilute (myosin heavy chain)
2217	EST01738	S	Gelation factor ABP-280
1885	EST01639	S	Histocompatibility antigen modifier 1
77	EST00257	S	Kinesin
SEQ ID	EST#	Group	Putative Identification
78	EST00258	S	Kinesin
2245	EST01748	S	Kinesin
313	EST00276	S	Lysosomal membrane glycoprotein 1 (LAMP-1)
223	EST00368	S	Microtubule-associated protein 1B
824	EST01865	S	Microtubule-associated protein 1B
2032	EST01683	S	Microtubule-associated protein 1B
2017	EST01678	S	Milk fat globule membrane protein
1567	EST02610	S	Neural cell adhesion molecule L1
506	EST01471	S	Neuraxin
2368	EST01389	S	Radial spoke protein 3
951	EST01960	S	Spectrin, beta
2089	EST01699	S	Sperm membrane protein
653	EST01512	S	Tubulin, alpha
311	EST00270	S	Tubulin, beta
594	EST01490	S	Tubulin, beta
757	EST01542	S	Tubulin, beta
1245	EST02274	S	Tubulin, beta
1589	EST02634	S	Tubulin, beta
1468	EST02505	S	Matrin 3

-70-

1371	EST02402	S	Talin
1701	EST00853	S	Unc-104

Group Key: M: Metabolic, R: Regulatory, S: Structural

Table 11: Thirteen-Class Functional Groupings of ESTs

<u>SEO ID</u>	<u>EST#</u>	<u>Group</u>	<u>Putative Identification</u>
208	EST00250	CS	60K filarial antigen
2320	EST01784	CS	60K filarial antigen
1965	EST01664	CS	Amyloid A4
2068	EST01694	CS	Amyloid A4
2408	EST00244	CS	Amyloid A4
1880	EST01634	CS	Axonal glycoprotein TAG-1
1885	EST01639	CS	Histocompatibility antigen modifier 1
313	EST00276	CS	Lysosomal membrane glycoprotein 1 (LAMP-1)
2017	EST01678	CS	Milk fat globule membrane protein
1567	EST02610	CS	Neural cell adhesion molecule L1
2368	EST01389	CS	Radial spoke protein 3
2089	EST01699	CS	Sperm membrane protein
1277	EST02306	DC	Bib protein
188	EST00256	DC	Enhancer of split
43	EST00371	DC	Maternal G10 protein
1704	EST01580	DC	Myeloid differentiation primary response gene MyD1
227	EST00259	DC	Notch/Xotch
952	EST01961	DC	Notch/Xotch
346	EST01828	DC	Orthodentical homeotic protein
1408	EST02442	DC	Seven in absentia
97	EST00289	EM	Aconitase
310	EST00377	EM	Fo ATPase beta subunit, mitochondrial
485	EST01466	KP	Calmodulin-dependent protein kinase, type II, beta
993	EST02007	KP	Kinase 5 protein
1069	EST02087	KP	Protein kinase C, zeta
1933	EST01650	KP	Protein phosphatase 2A beta subunit
202	EST00298	KP	Protein-tyrosine phosphatase LRP
1348	EST02378	KP	cAMP-dependent protein kinase inhibitor
2302	EST01779	OG	Discs-large tumor suppressor
2353	EST01806	OG	Prohibitin
1478	EST02515	OG	Rab5
300	EST00232	OG	Transforming protein (dbl)
37	EST00038	OG	ras p21-like small GTP-binding protein (smg GDS)
102	EST00248	OG	rho H12/ ARH12
1834	EST01620	OM	AMP deaminase, brain
691	EST00675	OM	Alcohol dehydrogenase
396	EST01443	OM	CDPdiacylglycerol-serine O-phosphatidyltransferase
2192	EST01257	OM	Diacylglycerol kinase, lymphocyte
1441	EST02477	OM	Diamine acetyltransferase
2289	EST01325	OM	Fatty acid synthase
1020	EST02034	OM	Glutaminase
2326	EST01791	OM	Inositol-1,4,5-trisphosphate 3-kinase
1427	EST02463	OM	Long-chain-fatty-acid-CoA ligase
2226	EST01744	OM	NAD(P)+ transhydrogenase (B-specific)
1566	EST02609	OM	Neutrophil oxidase factor
1681	EST01573	OM	Nucleoside diphosphate kinase

<u>SEQ ID</u>	<u>EST#</u>	<u>Group</u>	<u>Putative Identification</u>
2254	EST01751	OM	Phosphatidylinositol-4,5-bisphosphate phosphodiesterase
1654	EST01572	OM	Protochlorophyllide reductase
2073	EST01697	OM	Succinate dehydrogenase flavoprotein
2138	EST01715	OM	Succinate dehydrogenase flavoprotein
1771	EST01601	OM	Thiosulfate sulfurtransferase (rhodanese)
2173	EST01724	PI	Lon protease
2297	EST01775	PI	Prohormone cleavage enzyme
9	EST00376	PI	Prolyl endopeptidase
1726	EST01588	PI	XPR2 alkaline extracellular protease
1147	EST02169	PP	Tyrosine kinase
2282	EST01764	RT	Lamin B receptor
189	EST00282	RT	trkB
251	EST00370	SC	Actin, other
2146	EST01218	SC	Actin, other
248	EST00271	SC	Actinin, alpha
891	EST01891	SC	Actinin, alpha
1500	EST02538	SC	Actinin, alpha
132	EST00110	SC	Agrin
1852	EST01625	SC	Agrin
2004	EST01676	SC	Cofilin
650	EST00642	SC	Dilute (myosin heavy chain)
2217	EST01738	SC	Gelation factor ABP-280
77	EST00257	SC	Kinesin
78	EST00258	SC	Kinesin
2245	EST01748	SC	Kinesin
1468	EST02505	SC	Matrin 3
223	EST00368	SC	Microtubule-associated protein 1B
824	EST01865	SC	Microtubule-associated protein 1B
2032	EST01683	SC	Microtubule-associated protein 1B
506	EST01471	SC	Neuraxin
951	EST01960	SC	Spectrin, beta
1371	EST02402	SC	Talin
653	EST01512	SC	Tubulin, alpha
311	EST00270	SC	Tubulin, beta
594	EST01490	SC	Tubulin, beta
757	EST01542	SC	Tubulin, beta
1245	EST02274	SC	Tubulin, beta
1589	EST02634	SC	Tubulin, beta
1701	EST00853	SC	Unc-104
969	EST01982	ST	ADP-ribosylation factor 1
1126	EST02146	ST	Calbindin D28
1910	EST01645	ST	Calmodulin
161	EST00247	ST	MARCKS (myristoylated alanine-rich protein kinase)
769	EST00734	ST	MARCKS homolog
1386	EST02418	ST	MARCKS homolog
1931	EST01041	ST	cAMP-regulated phosphoprotein
1413	EST02447	ST	cAMP-specific phosphodiesterase
299	EST00249	ST	smg p25A GDP dissociation inhibitor

<u>SEQ ID</u>	<u>EST#</u>	<u>Group</u>	<u>Putative Identification</u>
2092	EST01700	TP	Anion exchanger homolog AE3
1956	EST01663	TP	Ca ²⁺ -transporting ATPase 2
1039	EST02055	TP	Calcium channel
1667	EST00825	TP	Gamma-aminobutyric acid transporter
1412	EST02446	TP	Glutamate-aspartate carrier protein
913	EST01913	TT	Clathrin coat assembly protein AP50 homolog
1035	EST02051	TT	J1 protein
93	EST00287	TT	Processing enhancing protein
38	EST00374	TT	RNA polymerase II 6th subunit (RPO26)
1715	EST01583	TT	Ribosomal protein L18a
1856	EST01627	TT	Ribosomal protein L1a
1974	EST01667	TT	Ribosomal protein L3
301	EST00300	TT	Ribosomal protein L30
22	EST00301	TT	Ribosomal protein S10
2402	EST01826	TT	Ribosomal protein S10
463	EST01459	TT	Ribosomal protein YL10
2121	EST01711	TT	Valine-tRNA ligase
1332	EST02362	TX	GA binding protein, beta subunit
1229	EST02258	TX	KUP protein
1395	EST02429	TX	Nuclear factor 1-like protein (NF1)
187	EST00152	TX	Wilm's tumor-related protein
249	EST00275	TX	Zinc Finger Proteins
413	EST01446	TX	Zinc Finger Proteins
469	EST01460	TX	Zinc Finger Proteins
833	EST01560	TX	Zinc Finger Proteins
1230	EST02259	TX	Zinc finger proteins
1496	EST02534	TX	Zinc finger proteins
2324	EST01352	TX	Zinc Finger Proteins

Group Key: CS: Cell Surface, DC: Developmental Control, EM: Energy Metabolism, KP: Kinases and Phosphatases, OG: Oncogenes, OM: Other Metabolism, PI, Peptidases and Peptidase Inhibitors, RT: Receptors, SC: Structural and Cytoskeletal, ST: Signal Transduction, TP: Transporters, TT: Transcription, Translation, and Subcellular Localization, TX: Transcription Factors.

-74-

EXAMPLE 11

cDNA Libraries Generated From Specific Genomic DNA
by Exon Expression & Amplification

5 Exon amplification was used to express potential exons from genomic DNA in a recombinant vector that contains some of the signals necessary for splicing. If an exon is present in the proper orientation in the vector, that exon will be
10 spliced in a mammalian cell and will become part of the mRNA of that cell. The exon splice-product can be purified from other mRNA in the cell by conversion of the mRNA to cDNA and selective amplification of the recombinant splice-product cDNAs. Cosmid DNA from human chromosome 19q13.3 was digested
15 with BamHI or BamHI/BglIII restriction enzymes. The fragments generated were collected and size specifically cloned into an expression vector (Buckler, et al. Proc. Nat'l. Acad. Sci. USA, 88:4005-4009 (1991)). After transfection by electroporation of these constructs into COS cells, RNA
20 transcripts were generated using the SV40 early promoter and a polyadenylation signal derived from SV40 both present in the expression vector. When a fragment of genomic DNA contains an entire exon with flanking intron sequence in the sense orientation, the exon should be retained in the mature poly(A)+ cytoplasmic RNA. Therefore, the mRNA was used as
25 template for cDNA synthesis using reverse transcriptase and vector-priming. Subsequently, the cDNAs were amplified by vector-priming using PCR. A fraction of this first PCR product was reamplified using internal vector-primers containing terminal cloning sites. These products were end-repaired with T4 DNA polymerase, digested with the appropriate restriction enzymes, gel purified and cloned into pBluescript vectors. The constructs were transfected into
30 XL1-Blue competent cells and plated on LB/X-gal/IPTG/ampicillin plates. White colonies were selected and expanded to prepare DNA templates as described in Example 2.

35

-75-

When multiple cosmids or YAC clones were used as the source DNA, a pool of specific expressed exons was obtained as a cDNA library. The EST/cDNAs sequenced from this specific library are disclosed herein as SEQ ID NOS: 2412-2417.

5

EXAMPLE 12**PCR Amplification from Predicted Exons**

10 Computational analyses can be applied to genomic DNA sequences to predict protein coding regions. The coding region prediction program CRM (E. Uberbacher and R. Mural, Proc. Natl. Acad. Sci. USA 88:11261-5 (1991)) finds open reading frames and classifies them according to their probability of being coding regions. These regions are subsequently examined using the GM program (C. Fields and C. Soderlund, Comp. Applic. Biosci. 6: 263, 1990), which predicts intron-exon structure. PCR primers are then designed to amplify the predicted exons and used to test human cDNA libraries (for example, fetal brain or placental libraries) for the presence of these putative exons using a PCR assay.

20 This strategy has been successfully applied in two large scale genomic sequencing projects, the Huntington's locus of human chromosome 4p16.3 (McCombie, et al., submitted) and human chromosome locus 19q13.3 (Martin-Gallardo, et al., submitted). Sequences from eleven predicted exons from chromosome 4 were present in tested cDNA libraries, indicating that this region has at least two and probably three expressed genes. In one case, the method resulted in an amplification product which spanned two predicted exons. (SEQ ID NO: 2411.) When sequenced, this PCR product indicated the presence of the two exons from which the primers were initially chosen, as well as an intervening exon which was also predicted by the CRM program, but not the intervening genomic sequences. In a similar fashion, the presence of the two predicted genes in the chromosome 19

35

-76-

sequence was confirmed by sequencing PCR products. SEQ ID NO 2410, includes a partial exon of one of these genes.

EXAMPLE 13

5 Complete Sequence of EST Clone Inserts

There are a number of methods known to those with skill in the art of molecular biology, to obtain sequence information from the cDNAs corresponding to the EST sequences. Procedures for these methods are provided in 10 Basic Methods in Molecular Biology (David et al. *supra*). One way to acquire more information about the cDNA from which an EST was derived is to sequence the remainder of the cDNA clone. The complete sequence of the inserts of four EST clones (representing SEQ ID NOs 188, 189, 223, and 227) was 15 determined using Exonuclease III deletions. Briefly, EST clones were digested with the restriction enzymes SalI and KpnI or PstI and BamHI (for deletions from the Forward primer and Reverse primer ends of the insert, respectively). The 20 KpnI and PstI enzymes leave 3' sticky ends following digestion, which Exonuclease III is unable to bind. This results in unidirectional deletions into the cDNA insert leaving the vector sequence undisturbed. After addition of Exonuclease III to the Forward and Reverse deletion 25 reactions, aliquots of the reaction were removed at defined time intervals and the reaction was stopped to prevent further deletion. S1 nuclease and Klenow DNA polymerase were added to create blunt ended fragments suitable for ligation.

Samples for each time point was purified by 30 electrophoresis through an agarose gel and religated. Two to four representative clones from each time point in each direction were sequenced to give between 200 and 400 base pairs of sequence data. Careful selection of deletion conditions and time points allow a deletion series of 35 approximately 100-200 base pairs difference in length at each consecutive time point. Sequence fragments were reassembled into a redundant contiguous sequence using the INHERIT

-77-

software from Applied Biosystems, Inc. (Foster City, CA). In this way, the complete insert from these four cDNA clones was sequenced on both strands to an average redundancy between three and four (each base was sequenced between three and four times, on average). Those complete insert sequences are disclosed herein as SEQ ID 2418, 2419, 2420, and 2421, corresponding to original ESTs with SEQ ID 223, 189, 227, and 188, respectively.

EXAMPLE 14

Determining Reading Frame, Orientation, Coding Regions: ESTs and Complete cDNA Sequences

Once the complete cDNA sequence has been determined in accordance with Example 13, the reading frame, orientation, and coding regions are determined by computer techniques. (The complete coding region is considered to be the largest open reading frame from a methionine to a stop codon.)

Specifically, the CRM program on the GRAIL server is used as explained in Example 9 to determine probable coding regions. This information is supplemented by location of start and stop codons. Where possible, the results of the CRM analysis are validated by comparison of the cDNA sequence to known sequences using database matching, in accordance with Examples 3 and 4. If a match of 50% (or even less) is found in any particular reading frame and orientation, this serves to verify corresponding CRM results. Alternatively, database matches can be used to determine reading frame and orientation without use of the CRM program. Of course, if the cDNA is derived from a directional library, the probable orientation is already known.

EXAMPLE 15

Preparation of PCR Primers and Amplification of DNA

The EST sequences and the corresponding cDNA sequences and genomic sequences may be used, in accordance with the

-78-

present invention, to prepare PCR primers for a variety of applications. The PCR primers are preferably at least 15 bases, and more preferably at least 18 bases in length. The procedure of Example 5 is repeated using the desired EST, or using the corresponding cDNA or genomic DNA sequence from Example 13. It is preferred that the primer pairs have approximately the same G/C ratio, so that melting temperatures are approximately the same. When screening cDNA, introns are of no concern; however, when screening genomic DNA, primers should be selected to avoid reading across introns, which usually are too large to amplify. The PCR primers and amplified DNA of this Example find use in the Examples that follow.

EXAMPLE 16

Forensic Matching by DNA Sequencing

In one exemplary method, DNA samples are isolated from forensic specimens of, for example, hair, semen, blood or skin cells by conventional methods. A panel of PCR primers derived from a number of the sequences of Example 1, 2, 11, 12 and/or 13 is then utilized in accordance with Example 12 to obtain DNA of approximately 100-200 bases in length from the forensic specimen. Corresponding sequences are obtained from a suspect. Each of these identification DNAs is then sequenced, and a simple database comparison determines the differences, if any, between the sequences from the suspect and those from the sample. Statistically significant differences between the suspect's DNA sequences and those from the sample conclusively prove a lack of identity. This lack of identity can be proven, for example, with only one sequence. Identity, on the other hand, should be demonstrated with a large number of sequences, all matching. Preferably, a minimum of 50 statistically identical sequences

-79-

of 100 bases in length are used to prove identity between the suspect and the sample.

EXAMPLE 17

5

Positive Identification by DNA Sequencing

The technique outlined in the previous example may also be used on a larger scale to provide a unique fingerprint-type identification of any individual. In this technique, primers are prepared from a large number of sequences from Examples 1, 2, 11, 12 and/or 13. Preferably, 20 to 50 different primers are used. These primers are used to obtain a corresponding number of PCR-generated DNA segments from the individual in question in accordance with Example 15. Each of these DNA segments is sequenced, using the methods set forth in Example 1. The database of sequences generated through this procedure uniquely identifies the individual from whom the sequences were obtained. The same panel of primers may then be used at any later time to absolutely correlate tissue or other biological specimen with that individual.

20

EXAMPLE 18

25

Southern Blot Forensic Identification

The procedure of Example 17 is repeated to obtain a panel of from 10 to 2000 amplified sequences from an individual and a specimen. This PCR-generated DNA is then digested with one or a combination of, preferably, four base specific restriction enzymes. Such enzymes are commercially available and known to those of skill in the art. After digestion, the resultant gene fragments are size separated in multiple duplicate wells on an agarose gel and transferred to nitrocellulose using Southern blotting techniques well known to those with skill in the art. For a review of Southern

30

35

-80-

blotting see Davis et al. (Basic Methods in Molecular Biology, 1986, Elsevier Press. pp 62-65).

A panel of ESTs or complete cDNA sequences from Examples 1, 2, and/or 13, or fragments thereof of at least 15 bases, are radioactively or colorimetrically labeled using end-labeled oligonucleotides derived from the ESTs, nick translated sequences or the like using methods known in the art and hybridized to the Southern blot using techniques known in the art (Davis et al., supra). Preferably, at least 5 to 10 of these labeled probes are used, and more preferably at least about 20 or 30 are used to provide a unique pattern. The resultant bands appearing from the hybridization of a large sample of ESTs will be a unique identifier. Since the restriction enzyme cleavage will be different for every individual, the band pattern on the Southern blot will also be unique. Increasing the number of EST probes will provide a statistically higher level of confidence in the identification since there will be an increased number of sets of bands used for identification.

EXAMPLE 19

Dot Blot Identification Procedure

Another technique for identifying individuals using the sequences disclosed herein utilizes a dot blot hybridization technique.

Genomic DNA is isolated from nuclei of subject to be identified. Oligonucleotide probes of approximately 30 bp in length were synthesized that correspond to sequences from the ESTs. The probes are used to hybridize to the genomic DNA through conditions known to those in the art. The oligonucleotides are end labelled with P^{32} using polynucleotide kinase (Pharmacia). Dot Blots are created by spotting about 50 ng cDNA of at least 10, preferably at least 50 sequences corresponding to a variety of the Sequence ID

-81-

5 NOs provided in Table 7 onto nitrocellulose or the like using
a vacuum dot blot manifold (BioRad, Richmond California).
The nitrocellulose filter containing the EST clone sequences
is baked or UV linked to the filter, prehybridized and
10 hybridized with labeled probe using techniques known in the
art (Davis et al. supra). The ³²P labeled DNA fragments are
sequentially hybridized with successively stringent
conditions to detect minimal differences between the 30 bp
sequence and the DNA. Tetramethylammonium chloride is useful
15 for identifying clones containing small numbers of nucleotide
mismatches (Wood et al., Proc. Natl. Acad. Sci. USA
82(6):1585-1588 (1985) which is hereby incorporated by
reference. A unique pattern of dots distinguishes one
individual from another individuals.

EXAMPLE 20

Alternative "Fingerprint" Identification Technique

20 EST sequences and the corresponding complete cDNA
sequences can be used to create a unique fingerprint for an
individual. Thus pools of EST sequences can be used in
forensics, paternity suits or the like to differentiate one
individual from another.

25 Entire EST sequences can be used; similarly
oligonucleotides can be prepared from EST sequences. In this
example, 20-mer oligonucleotides are prepared from 200 EST
sequences using commercially available oligonucleotide
services such as Oligos Etc., Wilsonville, OR. Patient cell
30 samples are processed for DNA using techniques well known to
those with skill in the art. The nucleic acid is digested
with restriction enzymes EcoRI and XbaI. Following
digestion, samples are applied to wells for electrophoresis.
The procedure, as known in the art, may be modified to
35 accommodate polyacrylamide electrophoresis, however in this
example, samples containing 5 ug of DNA are loaded into wells

-82-

and separated on 0.8% agarose gels. The gels are transferred using Southern blotting techniques onto nitrocellulose.

10 ng of each of the oligos are pooled and end-labeled with P³². The nitrocellulose is prehybridized with blocking solution and hybridized with the labeled probes. Following hybridization and washing, the nitrocellulose filter is exposed to X-Omat AR X-ray film. The resulting hybridization pattern will be unique for each individual.

It is additionally contemplated within this example that the representative number of EST sequences can be varied for additional accuracy or clarity.

EXAMPLE 21

Identification of genes associated with hereditary diseases

This example illustrates an approach useful for the association of EST sequences with particular phenotypic characteristics. In this example, a particular EST is used as a test probe to associate that EST with a particular phenotypic characteristic.

An EST clone corresponding to EST01643, (SEQ ID NO 1894) maps to a gene rich region of chromosome 6. EST clone HHCMH89, from which EST01643 was derived, was mapped to chromosome 6p21 by Dr. Julie Korenberg of UCLA/Cedar Sinai Hospital using FISH. A search of Mendelian Inheritance in Man (supra) revealed 6p21 to be a very gene rich region containing several known genes and several diseases for which genes have not been identified. The cDNA encoded by EST clone HHCMH89 thus becomes an immediate candidate for each of these genetic diseases.

Cells from patients with these diseases are isolated and expanded in culture. PCR primers from the EST sequences are used to screen genomic DNA and RNA or cDNA from the patients. ESTs that are not amplified in the patients can be positively associated with a particular disease by further analysis.

EXAMPLE 22

Identification of a gene associated with
Angelman's disease

5

Angelman's disease (AD) is characterized by deletions on the long arm of chromosome 15 (15q11q13) (Williams et al. Am. J. Med. Genet. 32:339-345 (1989) hereby incorporated by reference). The symptoms of the disease include developmental delay, seizures, inappropriate laughter and ataxic movements. These symptoms suggest that the disorder is a neurologic deficiency. This prophetic example illustrates how ESTs, preferably obtained from a cDNA library from human brain, may be used in identifying the defective gene or genes associated with Angelman's Disease. (The example is based on analogous work with genomic DNA, rather than cDNA and ESTs, in identifying the genetic defect associated with Angelman's Disease.) This example also illustrates how EST sequences may generally be used for identifying gene sequences associated with an inherited disease that is mapped to a chromosome location.

ESTs are screened using techniques described in Example 5 and Example 7 to identify those ESTs that localize to the long arm of chromosome 15 and preferably localize to chromosome 15 bands 15q11q13 from normal patients. ESTs that bind to the long arm of chromosome 15 are hybridized to chromosome 15 from AD patients. These studies are preferably performed using either fluorescence in situ hybridization or using somatic cell hybrids that contain fragments from the long arm of chromosome 15 from AD patients. Those chromosome 15-specific ESTs that do not map to chromosome 15 from AD patients are useful as markers for Angelman's Disease and can be incorporated into diagnostics for genetic screening. These ESTs are associated with chromosome deletions present in Angelman's disease. Identification of the gene associated with these AD negative ESTs and an analysis of the polypeptides encoded by the genes

-84-

from normal patients is essential for providing gene or other therapies for AD patients.

Genetic diseases are not always accompanied by gene deletions. Therefore, it is also important to use the ESTs that bind to bands 15q11q13 from AD patients as tools to identify the polymorphisms present within the disease population. Restriction fragment length polymorphism (RFLP) analysis can be performed on patient cells from AD disease or from somatic cell hybrids created using the long arm of chromosome 15. For a review of RFLP techniques see Donis-Keller et al. (Cell 51:319-337 (1987) hereby incorporated by reference). DNA is isolated from the somatic cell lines or from cells from AD patients. The DNA is digested with one or more restriction enzymes according to techniques of Donis-Keller et al. The resulting fragments are separated by gel electrophoresis, denatured, transferred to nitrocellulose and hybridized with the selected radio-labeled ESTs that localize to the region of interest. The autoradiographic pattern is compared both to a number of AD patients and to normal patients. Common patterns of EST hybridization in AD patients that are not present in normal patients indicates that the genes associated with these ESTs are candidate genes affected by AD.

cDNA libraries are prepared from the somatic cell hybrids from AD patients. Libraries are prepared using Lambda Zap II Library Kits (Stratagene, La Jolla, California) or other commercially available library kits. The ESTs of interest are used as probes to identify those bacterial colonies carrying genes corresponding to the EST probes. Positive clones are sequenced and the sequences are compared to homologous gene sequences derived from normal patients.

Alterations, including deletions and substitutions, within gene sequences, associated with bands 15q11q13, are thus positively identified and associated with AD disease. Wagstaff et al. were able to identify deletions and substitutions in sequences encoding the GABA_A receptor

-85-

protein subunit from patients with Angelman's disease (Am. J. Hum. Genet. 49:330-337, (1991)). It is likely that other genes will additionally be associated with the disease.

5

EXAMPLE 23**Preparation and Use of Antisense Oligonucleotides**

Antisense RNA molecules are known to be useful for
10 regulating translation within the cell. Antisense RNA
molecules can be produced from EST sequences or from the
corresponding gene sequences. These antisense molecules can
be used as diagnostic probes to determine whether or not a
particular gene is expressed in a cell. Similarly, the
15 antisense molecules can be used as a therapeutic to regulate
gene expression once the EST is associated with a particular
disease (see Example 22).

The antisense molecules are obtained from a nucleotide
sequence by reversing the orientation of the coding region
20 with regard to the promoter. Thus, the antisense RNA is
complementary to the corresponding mRNA. For a review of
antisense design see Green et al., Ann. Rev. Biochem. 55:569-
597 (1986), which is hereby incorporated by reference. The
antisense sequences can contain modified sugar phosphate
25 backbones to increase stability and make them less sensitive
to RNase activity. Examples of the modifications are
described by Rossi et al., Pharmacol. Ther. 50(2):245-254,
(1991).

Antisense molecules are introduced into cells that
30 express the gene corresponding to the EST of interest in
culture. In a preferred application of this invention, the
polypeptide encoded by the gene is first identified, so that
the effectiveness of antisense inhibition on translation can
be monitored using techniques that include but are not
35 limited to antibody-mediated tests such as RIAs and ELISA,
functional assays, or radiolabelling. The antisense molecule
is introduced into the cells by diffusion or by transfection

-86-

procedures known in the art. The molecules are introduced onto cell samples at a number of different concentrations preferably between $1 \times 10^{-10} \text{M}$ to $1 \times 10^{-4} \text{M}$. Once the minimum concentration that can adequately control translation is identified, the optimized dose is translated into a dosage suitable for use in vivo. For example, an inhibiting concentration in culture of 1×10^{-7} translates into a dose of approximately 0.6 mg/kg bodyweight. Levels of oligonucleotide approaching 100 mg/kg bodyweight or higher may be possible after testing the toxicity of the oligonucleotide in laboratory animals.

The antisense can be introduced into the body as a bare or naked oligonucleotide, oligonucleotide encapsulated in lipid, oligonucleotide sequence encapsidated by viral protein, or as oligonucleotide contained in an expression vector such as those described in Example 25. The antisense oligonucleotide is preferably introduced into the vertebrate by injection. It is additionally contemplated that cells from the vertebrate are removed, treated with the antisense oligonucleotide, and reintroduced into the vertebrate. It is further contemplated that the antisense oligonucleotide sequence is incorporated into a ribozyme sequence to enable the antisense to bind and cleave its target. For technical applications of ribozyme and antisense oligonucleotides see Rossi et al.

EXAMPLE 24

Preparation and use of Triple Helix Probes

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity as it is associated with a particular gene. The EST sequences or complete sequences of the present invention or, more preferably, a portion of those sequences, can be used to inhibit gene

-87-

expression in individuals having diseases associated with a particular gene. Similarly, a portion of the EST or corresponding gene sequence can be used to study the effect of inhibiting transcription of a particular gene within a cell. Traditionally, homopurine sequences were considered the most useful. However, homopyrimidine sequences can also inhibit gene expression. Thus, both types of sequences from either the EST or from the gene corresponding to the EST are contemplated within the scope of this invention. Homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. As an example, 10-mer to 20-mer homopyrimidine sequences from the ESTs can be used to inhibit expression from homopurine sequences. SEQ ID NOS such as 282, 888, 719, 670, 994, 240, 873 and 761 contain homopyrimidine 15-mers. Moreover the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al. (Science 245:967-971 (1989), which is hereby incorporated by this reference).

The oligonucleotides may be prepared on an oligonucleotide synthesizer or they may be purchased commercially from a company specializing in custom oligonucleotide synthesis. The sequences are introduced into cells in culture using techniques known in the art that include but are not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake. Treated cells are monitored for altered cell function. These cell functions are predicted based upon the homologies of the gene, corresponding to the EST from which the oligonucleotide was derived, with known genes sequences that have been associated

-88-

with a particular function. The cell functions can also be predicted based on the presence of abnormal physiologies within cells derived from individuals with a particular inherited disease, particularly when the EST is associated with the disease using techniques described in Example 22.

EXAMPLE 25

Gene expression from DNA Sequences Corresponding to ESTs

10

A gene sequence of the present invention coding for all or part of a human gene product is introduced into an expression vector using conventional technology. (Techniques to transfer cloned sequences into expression vectors that direct protein translation in mammalian, yeast, insect or bacterial expression systems are well known in the art.) Commercially available vectors and expression systems are available from a variety of suppliers including Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence may be optimized for the particular expression organism, as explained by Hatfield, et al., U.S. Patent No. 5,082,767, incorporated herein by this reference.

25

The following is provided as one exemplary method to generate polypeptide from cloned cDNA sequences. The cDNA from the EST of interest is sequenced to identify the methionine initiation codon for the gene and the poly A sequence. If the cDNA lacks a poly A sequence, this sequence can be added to the construct by, for example, splicing out the Poly A sequence from pSG5 (Stratagene) using BglI and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene). pXT1 contains the LTRs and a portion of the gag gene from Moloney Murine Leukemia Virus. The position of the LTRs in the construct allow efficient stable transfection. The vector

30

35

-89-

includes the Herpes Simplex Thymidine Kinase promoter and the selectable neomycin gene. The cDNA is obtained by PCR from the bacterial vector using oligonucleotide primers complementary to the cDNA and containing restriction endonuclease sequences for Pst I incorporated into the 5' primer and BglII at the 5' end of the corresponding cDNA 3' primer, taking care to ensure that the cDNA is positioned inframe with the poly A sequence. The purified fragment obtained from the resulting PCR reaction is digested with PstI, blunt ended with an exonuclease, digested with Bgl II, purified and ligated to pXT1, now containing a poly A sequence and digested BglII.

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri). The protein is preferrably released into the supernatant. However if the protein has membrane binding domains, the protein may additionally be retained within the cell or expression may be restricted to the cell surface.

Since it may be necessary to purify and locate the transfected product, synthetic 15-mer peptides synthesized from the predicted cDNA sequence are injected into mice to generate antibody to the polypeptide encoded by the cDNA.

If antibody production is not possible, the cDNA sequence is additionally incorporated into eukaryotic expression vectors and expressed as a chimeric with, for example, β -globin. Antibody to β -globin is used to purify the chimeric. Corresponding protease cleavage sites engineered between the β -globin gene and the cDNA are then used to separate the two polypeptide fragments from one another after translation. One useful expression vector for generating β -globin chimerics is pSG5 (Stratagene). This vector encodes rabbit β -globin. Intron II of the rabbit β -globin gene facilitates splicing of the expressed transcript,

-90-

and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques as described are well known to those skilled in the art of molecular biology. Standard methods are published in methods texts such as Davis et al. and many of the methods are available from the technical assistance representatives from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from either construct using in vitro translation systems such as In vitro Express™ Translation Kit (Stratagene).

Example 26

Production of an Antibody to a Human Protein

Substantially pure protein or polypeptide is isolated from the transfected or transformed cells as described in Example 25. Concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibody to the protein can then be prepared as follows:

A. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes of any of the peptides identified and isolated as described can be prepared from murine hybridomas according to the classical method of Kohler, G. and Milstein, C., Nature 256:495 (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a

-91-

microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as Elisa, as originally described by Engvall, E., *Meth. Enzymol.* 70:419 (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al. *Basic Methods in Molecular Biology* Elsevier, New York. Section 21-2.

B. Polyclonal Antibody Production by Immunization

Polyclonal antiserum containing antibodies to heterogenous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein described above, which can be unmodified or modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. For example, small molecules tend to be less immunogenic than other and may require the use of carriers and adjuvant. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. *J. Clin. Endocrinol. Metab.* 33:988-991 (1971).

Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony, O. et al., Chap. 19 in: *Handbook of Experimental Immunology* D. Wier (ed) Blackwell (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum (about 12 μ M). Affinity of the antisera for the antigen is

-92-

determined by preparing competitive binding curves, as described, for example, by Fisher, D., Chap. 42 in: **Manual of Clinical Immunology**, 2d Ed. (Rose and Friedman, eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

5 Antibody preparations prepared according to either protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a
10 biological sample.

EXAMPLE 27

Identification of Tissue Types or Cell Species by Means of Labeled Tissue Specific Antibodies

15 Identification of specific tissues is accomplished by the visualization of tissue specific antigens by means of antibody preparations according to Example 26 which are conjugated, directly or indirectly to a detectable marker.
20 Selected labeled antibody species bind to their specific antigen binding partner in tissue sections, cell suspensions, or in extracts of soluble proteins from a tissue sample to provide a pattern for qualitative or semi-qualitative interpretation.

25 Antisera for these procedures must have a potency exceeding that of the native preparation, and for that reason, antibodies are concentrated to a mg/ml level by isolation of the gamma globulin fraction, for example, by ion-exchange chromatography or by ammonium sulfate
30 fractionation. Also, to provide the most specific antisera, unwanted antibodies, for example to common proteins, must be removed from the gamma globulin fraction, for example by means of insoluble immunoabsorbents, before the antibodies are labeled with the marker. Either monoclonal or
35 heterologous antisera is suitable for either procedure.

A. Immunohistochemical Techniques

Purified, high-titer antibodies, prepared as described above, are conjugated to a detectable marker, as described, for example, by Fudenberg, H., Chap. 26 in: **Basic & Clinical Immunology**, 3rd Ed. Lange, Los Altos, California (1980) or Rose, N. et al., Chap. 12 in: **Methods in Immunodiagnosis**, 2d Ed. John Wiley & Sons, New York (1980).

A fluorescent marker, either fluorescein or rhodamine, is preferred, but antibodies can also be labeled with an enzyme that supports a color producing reaction with a substrate, such as horseradish peroxidase. Markers can be added to tissue-bound antibody in a second step, as described below. Alternatively, the specific antitissue antibodies can be labeled with ferritin or other electron dense particles, and localization of the ferritin coupled antigen-antibody complexes achieved by means of an electron microscope. In yet another approach, the antibodies are radiolabeled, with, for example ^{125}I , and detected by overlaying the antibody treated preparation with photographic emulsion.

Preparations to carry out the procedures can comprise monoclonal or polyclonal antibodies to a single gene copy or protein, identified as specific to a tissue type, for example, brain tissue, or antibody preparations to several antigenically distinct tissue specific antigens can be used in panels, independently or in mixtures, as required.

Tissue sections and cell suspensions are prepared for immunohistochemical examination according to common histological techniques. Multiple cryostat sections (about 4 μm , unfixed) of the unknown tissue and known control, are mounted and each slide covered with different dilutions of the antibody preparation. Sections of known and unknown tissues should also be treated with preparations to provide a positive control, a negative control, for example, pre-immune sera, and a control for non-specific staining, for example, buffer.

-94-

Treated sections are incubated in a humid chamber for 30 min at room temperature, rinsed, then washed in buffer for 30-45 min. Excess fluid is blotted away, and the marker developed.

5 If the tissue specific antibody was not labeled in the first incubation, it can be labeled at this time in a second antibody-antibody reaction, for example, by adding fluorescein- or enzyme-conjugated antibody against the immunoglobulin class of the antiserum-producing species, for
10 example, fluorescein labeled antibody to mouse IgG. Such labeled sera are commercially available.

15 The antigen found in the tissues by the above procedure can be quantified by measuring the intensity of color or fluorescence on the tissue section, and calibrating that signal using appropriate standards.

B. Identification of Tissue Specific Soluble Proteins

20 The visualization of tissue specific proteins and identification of unknown tissues from that procedure is carried out using the labeled antibody reagents and detection strategy as described for immunohistochemistry; however the sample is prepared according to an electrophoretic technique to distribute the proteins extracted from the tissue in an orderly array on the basis of molecular weight for detection.

25 A tissue sample is homogenized using a Virtis apparatus; cell suspensions are disrupted by Dounce homogenization or osmotic lysis, using detergents in either case as required to disrupt cell membranes, as is the practice in the art. Insoluble cell components such as nuclei, microsomes, and membrane fragments are removed by ultracentrifugation, and
30 the soluble protein-containing fraction concentrated if necessary and reserved for analysis.

35 A sample of the soluble protein solution is resolved into individual protein species by conventional SDS polyacrylamide electrophoresis as described, for example, by Davis, L. et al., Section 19-2 in: *Basic Methods in Molecular Biology* (P. Leder, ed), Elsevier, New York (1986), using a

-95-

range of amounts of polyacrylamide in a set of gels to resolve the entire molecular weight range of proteins to be detected in the sample. A size marker is run in parallel for purposes of estimating molecular weights of the constituent proteins. Sample size for analysis is a convenient volume of from 5-50 μ l, and containing from about 1 to 100 μ g protein. An aliquot of each of the resolved proteins is transferred by blotting to a nitrocellulose filter paper, a process that maintains the pattern of resolution. Multiple copies are prepared. The procedure, known as Western Blot Analysis, is well described in Davis, L. et al., (above) Section 19-3. One set of nitrocellulose blots is stained with Coomassie Blue dye to visualize the entire set of proteins for comparison with the antibody bound proteins. The remaining nitrocellulose filters are then incubated with a solution of one or more specific antisera to tissue specific proteins prepared as described in Example 26. In this procedure, as in procedure A above, appropriate positive and negative sample and reagent controls are run.

In either procedure A or B, a detectable label can be attached to the primary tissue antigen-primary antibody complex according to various strategies and permutations thereof. In a straightforward approach, the primary specific antibody can be labeled; alternatively, the unlabeled complex can be bound by a labeled secondary anti-IgG antibody. In other approaches, either the primary or secondary antibody is conjugated to a biotin molecule, which can, in a subsequent step, bind an avidin conjugated marker. According to yet another strategy, enzyme labeled or radioactive protein A, which has the property of binding to any IgG, is bound in a final step to either the primary or secondary antibody.

The visualization of tissue specific antigen binding at levels above those seen in control tissues to one or more tissue specific antibodies, prepared from the gene sequences identified from EST sequences, can identify tissues of unknown origin, for example, forensic samples, or

-96-

differentiated tumor tissue that has metastasized to foreign bodily sites.

The entire contents of all references cited above are hereby incorporated by reference.

5 While the present invention has been described in some detail for purposes of clarity and understanding, one skilled in the art will appreciate that various changes in form and detail can be made without departing from the true scope of the invention.

10

VII. Correlation of EST and Clone Identifiers

15 The EST sequences of the present invention are identified herein by SEQ ID NO, and are identified in the GenBank database by a different number, are identified in the inventors' lab (and upcoming publications) by EST number, and clones have been submitted to the American Type Culture Collection (Rockville, Maryland USA) under clone names. Table 12 cross references those different numbers for the ESTs from cDNA, SEQ ID NOS 1-2409.

20 Certain Sequence ID NOS are excluded from some claims based on their homology to known non-human sequences (See Table 2).

Table 12. SEQ ID NO Cross References

SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone
1	EST00007	M61959	HFBAD1	64	EST000066	M62010	HHC13	128	EST000252	M62191	HHC57	129	EST000321	M62254	HHC60	130	EST000355	M62283	HHC61
2	EST00009	M61953	HFBAD5	65	EST000067	M62011	HHC18	129	EST000322	M62255	HHC62	131	EST000356	M62284	HHC63	132	EST000357	M62285	HHC64
3	EST00010	M61961	HFBAD7	66	EST000068	M62012	HHC21	130	EST000358	M62286	HHC65	133	EST000359	M62287	HHC66	134	EST000360	M62288	HHC67
4	EST00011	M61962	HFBAD8	67	EST000069	M62013	HHC22	131	EST000361	M62289	HHC66	135	EST000362	M62290	HHC67	136	EST000363	M62291	HHC68
5	EST00012	M61963	HFBAD9	68	EST000070	M62014	HHC23	132	EST000364	M62292	HHC69	137	EST000365	M62293	HHC70	138	EST000366	M62294	HHC71
6	EST00013	M61964	HFBAD10	69	EST000071	M62015	HHC24	133	EST000367	M62295	HHC70	139	EST000368	M62296	HHC71	139	EST000369	M62297	HHC72
7	EST00014	M61965	HFBAD11	70	EST000072	M62016	HHC25	134	EST000370	M62298	HHC71	140	EST000371	M62299	HHC72	140	EST000372	M62300	HHC73
8	EST00015	M61966	HFBAD12	71	EST000073	M62017	HHC26	135	EST000373	M62301	HHC72	141	EST000374	M62302	HHC73	141	EST000375	M62303	HHC74
9	EST00016	M61967	HFBAD13	72	EST000074	M62018	HHC27	136	EST000376	M62304	HHC73	142	EST000377	M62305	HHC74	142	EST000378	M62306	HHC75
10	EST00017	M61968	HFBAD14	73	EST000075	M62019	HHC28	137	EST000379	M62307	HHC74	143	EST000380	M62308	HHC75	143	EST000381	M62309	HHC76
11	EST00018	M61969	HFBAD15	74	EST000076	M62020	HHC29	138	EST000382	M62310	HHC75	144	EST000383	M62311	HHC76	144	EST000384	M62312	HHC77
12	EST00019	M61970	HFBAD16	75	EST000077	M62021	HHC30	139	EST000385	M62313	HHC76	145	EST000386	M62314	HHC77	145	EST000387	M62315	HHC78
13	EST00020	M61971	HFBAD17	76	EST000078	M62022	HHC31	140	EST000388	M62316	HHC77	146	EST000389	M62317	HHC78	146	EST000390	M62318	HHC79
14	EST00021	M61972	HFBAD18	77	EST000079	M62023	HHC32	141	EST000391	M62319	HHC78	147	EST000392	M62320	HHC79	147	EST000393	M62321	HHC80
15	EST00022	M61973	HFBAD19	78	EST000080	M62024	HHC33	142	EST000394	M62322	HHC79	148	EST000395	M62323	HHC80	148	EST000396	M62324	HHC81
16	EST00023	M61974	HFBAD20	79	EST000081	M62025	HHC34	143	EST000397	M62325	HHC80	149	EST000398	M62326	HHC81	149	EST000399	M62327	HHC82
17	EST00024	M61975	HFBAD21	80	EST000082	M62026	HHC35	144	EST000399	M62328	HHC81	150	EST000400	M62329	HHC82	150	EST000401	M62330	HHC83
18	EST00025	M61976	HFBAD22	81	EST000083	M62027	HHC36	145	EST000402	M62331	HHC82	151	EST000403	M62332	HHC83	151	EST000404	M62333	HHC84
19	EST00026	M61977	HFBAD23	82	EST000084	M62028	HHC37	146	EST000405	M62334	HHC83	152	EST000406	M62335	HHC84	152	EST000407	M62336	HHC85
20	EST00027	M61978	HFBAD24	83	EST000085	M62029	HHC38	147	EST000408	M62337	HHC84	153	EST000409	M62338	HHC85	153	EST000410	M62339	HHC86
21	EST00028	M61979	HFBAD25	84	EST000086	M62030	HHC39	148	EST000411	M62340	HHC85	154	EST000412	M62341	HHC86	154	EST000413	M62342	HHC87
22	EST00029	M61980	HFBAD26	85	EST000087	M62031	HHC40	149	EST000414	M62343	HHC86	155	EST000415	M62344	HHC87	155	EST000416	M62345	HHC88
23	EST00030	M61981	HFBAD27	86	EST000088	M62032	HHC41	150	EST000417	M62346	HHC87	156	EST000418	M62347	HHC88	156	EST000419	M62348	HHC89
24	EST00031	M61982	HFBAD28	87	EST000089	M62033	HHC42	151	EST000420	M62349	HHC88	157	EST000421	M62350	HHC89	157	EST000422	M62351	HHC90
25	EST00032	M61983	HFBAD29	88	EST000090	M62034	HHC43	152	EST000423	M62352	HHC89	158	EST000424	M62353	HHC90	158	EST000425	M62354	HHC91
26	EST00033	M61984	HFBAD30	89	EST000091	M62035	HHC44	153	EST000426	M62355	HHC90	159	EST000427	M62356	HHC91	159	EST000428	M62357	HHC92
27	EST00034	M61985	HFBAD31	90	EST000092	M62036	HHC45	154	EST000429	M62358	HHC91	160	EST000430	M62359	HHC92	160	EST000431	M62360	HHC93
28	EST00035	M61986	HFBAD32	91	EST000093	M62037	HHC46	155	EST000432	M62361	HHC92	161	EST000433	M62362	HHC93	161	EST000434	M62363	HHC94
29	EST00036	M61987	HFBAD33	92	EST000094	M62038	HHC47	156	EST000435	M62364	HHC93	162	EST000436	M62365	HHC94	162	EST000437	M62366	HHC95
30	EST00037	M61988	HFBAD34	93	EST000095	M62039	HHC48	157	EST000438	M62367	HHC94	163	EST000439	M62368	HHC95	163	EST000440	M62369	HHC96
31	EST00038	M61989	HFBAD35	94	EST000096	M62040	HHC49	158	EST000441	M62370	HHC95	164	EST000442	M62371	HHC96	164	EST000443	M62372	HHC97
32	EST00039	M61990	HFBAD36	95	EST000097	M62041	HHC50	159	EST000444	M62373	HHC96	165	EST000445	M62374	HHC97	165	EST000446	M62375	HHC98
33	EST00040	M61991	HFBAD37	96	EST000098	M62042	HHC51	160	EST000447	M62376	HHC97	166	EST000448	M62377	HHC98	166	EST000449	M62378	HHC99
34	EST00041	M61992	HFBAD38	97	EST000099	M62043	HHC52	161	EST000450	M62379	HHC98	167	EST000451	M62380	HHC99	167	EST000452	M62381	HHC100
35	EST00042	M61993	HFBAD39	98	EST000100	M62044	HHC53	162	EST000453	M62382	HHC99	168	EST000454	M62383	HHC100	168	EST000455	M62384	HHC101
36	EST00043	M61994	HFBAD40	99	EST000101	M62045	HHC54	163	EST000456	M62385	HHC100	169	EST000457	M62386	HHC101	169	EST000458	M62387	HHC102
37	EST00044	M61995	HFBAD41	100	EST000102	M62046	HHC55	164	EST000459	M62388	HHC101	170	EST000460	M62389	HHC102	170	EST000461	M62390	HHC103
38	EST00045	M61996	HFBAD42	101	EST000103	M62047	HHC56	165	EST000462	M62391	HHC102	171	EST000463	M62392	HHC103	171	EST000464	M62393	HHC104
39	EST00046	M61997	HFBAD43	102	EST000104	M62048	HHC57	166	EST000465	M62394	HHC103	172	EST000466	M62395	HHC104	172	EST000467	M62396	HHC105
40	EST00047	M61998	HFBAD44	103	EST000105	M62049	HHC58	167	EST000468	M62397	HHC104	173	EST000469	M62398	HHC105	173	EST000470	M62399	HHC106
41	EST00048	M61999	HFBAD45	104	EST000106	M62050	HHC59	168	EST000471	M62400	HHC105	174	EST000472	M62401	HHC106	174	EST000473	M62402	HHC107
42	EST00049	M62000	HFBAD46	105	EST000107	M62051	HHC60	169	EST000474	M62403	HHC106	175	EST000475	M62404	HHC107	175	EST000476	M62405	HHC108
43	EST00050	M62001	HFBAD47	106	EST000108	M62052	HHC61	170	EST000477	M62406	HHC107	176	EST000478	M62407	HHC108	176	EST000479	M62408	HHC109
44	EST00051	M62002	HFBAD48	107	EST000109	M62053	HHC62	171	EST000480	M62409	HHC108	177	EST000481	M62410	HHC109	177	EST000482	M62411	HHC110
45	EST00052	M62003	HFBAD49	108	EST000110	M62054	HHC63	172	EST000483	M62412	HHC109	178	EST000484	M62413	HHC110	178	EST000485	M62414	HHC111
46	EST00053	M62004	HFBAD50	109	EST000111	M62055	HHC64	173	EST000486	M62415	HHC110	179	EST000487	M62416	HHC111	179	EST000488	M62417	HHC112
47	EST00054	M62005	HFBAD51	110	EST000112	M62056	HHC65	174	EST000489	M62418	HHC111	180	EST000490	M62419	HHC112	180	EST000491	M62420	HHC113
48	EST00055	M62006	HFBAD52	111	EST000113	M62057	HHC66	175	EST000492	M62421	HHC112	181	EST000493	M62422	HHC113	181	EST000494	M62423	HHC114
49	EST00056	M62007	HFBAD53	112	EST000114	M62058	HHC67	176	EST000495	M62424	HHC113								
50	EST00057	M62008	HFBAD54	113	EST000115	M62059	HHC68	177	EST000496	M62425	HHC114								
51	EST00058	M62009	HFBAD55	114	EST000116	M62060	HHC69	178	EST000497	M62426	HHC115								
52	EST00059	M62010	HFBAD56	115	EST000117	M62061	HHC70	179	EST000498	M62427	HHC116								
53	EST00060	M62011	HFBAD57	116	EST000118	M62062	HHC71	180	EST000499	M62428	HHC117								
54	EST00061	M62012	HFBAD58	117	EST000119	M62063	HHC72												
55	EST00062	M62013	HFBAD59	118	EST000120	M62064	HHC73												
56	EST00063	M62014	HFBAD60	119	EST000121	M62065	HHC74												
57	EST00064	M62015	HFBAD61	120	EST000122	M62066	HHC75												
58	EST00065	M62016	HFBAD62	121	EST000123	M62067	HHC76												
59	EST00066	M62017	HFBAD63	122	EST000124	M62068	HHC77												
60	EST00067	M62018	HFBAD64	123	EST000125	M62069	HHC78												
61	EST00068	M62019	HFBAD65	124	EST000126	M62070	HHC79												
62	EST00069	M62020	HFBAD66	125	EST000127	M62071	HHC80												
63	EST00070	M62021	HFBAD67	126	EST000128	M62072	HHC81												

SUBSTITUTE SHEET

-98-

SEQ. ID	EST#	GB#	Clone	SEQ. ID	EST#	GB#	Clone	SEQ. ID	EST#	GB#	Clone
182	EST003329	M62261	HHC159	317	EST003379	M62213	HHC586	373	EST01832	M85324	HF88A28
183	EST00148	M62089	HHC161	318	EST003380	M62136	HHC587	374	EST01833	M85325	HF88A27
184	EST00149	M62090	HHC162	319	EST003381	M62296	HHC589	375	EST01834	M85326	HF88A26
185	EST00150	M62091	HHC173	320	EST003382	M62137	HHC592	376	EST01835	M85327	HF88A25
186	EST00151	M62092	HHC174	321	EST003383	M62138	HHC595	377	EST01836	M85328	HF88A24
187	EST00152	M62093	HHC175	322	EST003384	M62139	HHC595	378	EST01837	M85329	HF88A23
188	EST00256	M62195	HHC179	323	EST003385	M62140	HHC598	379	EST01838	M85330	HF88A22
189	EST00282	M62219	HHC184	324	EST003386	M62267	HHC599	380	EST01839	M85331	HF88A21
190	EST00153	M62094	HHC185	325	EST003387	M62268	HHC599	381	EST01840	M85332	HF88A20
191	EST00154	M62095	HHC186	326	EST003388	M62269	HHC599	382	EST01841	M85333	HF88A19
192	EST00155	M62096	HHC188	327	EST003389	M62270	HHC599	383	EST01842	M85334	HF88A18
193	EST00156	M62097	HHC192	328	EST003390	M62271	HHC599	384	EST01843	M85335	HF88A17
194	EST00157	M62098	HHC193	329	EST003391	M62272	HHC599	385	EST01844	M85336	HF88A16
195	EST00158	M62099	HHC194	330	EST003392	M62273	HHC599	386	EST01845	M85337	HF88A15
196	EST00159	M62100	HHC195	331	EST003393	M62274	HHC599	387	EST01846	M85338	HF88A14
197	EST00160	M62101	HHC197	332	EST003394	M62275	HHC599	388	EST01847	M85339	HF88A13
198	EST00161	M62102	HHC199	333	EST003395	M62276	HHC599	389	EST01848	M85340	HF88A12
199	EST00162	M62103	HHC199	334	EST003396	M62277	HHC599	390	EST01849	M85341	HF88A11
200	EST00163	M62104	HHC199	335	EST003397	M62278	HHC599	391	EST01850	M85342	HF88A10
201	EST00164	M62105	HHC199	336	EST003398	M62279	HHC599	392	EST01851	M85343	HF88A09
202	EST00165	M62106	HHC199	337	EST003399	M62280	HHC599	393	EST01852	M85344	HF88A08
203	EST00166	M62107	HHC199	338	EST00400	M62281	HHC599	394	EST01853	M85345	HF88A07
204	EST00167	M62108	HHC199	339	EST00401	M62282	HHC599	395	EST01854	M85346	HF88A06
205	EST00168	M62109	HHC199	340	EST00402	M62283	HHC599	396	EST01855	M85347	HF88A05
206	EST00169	M62110	HHC199	341	EST00403	M62284	HHC599	397	EST01856	M85348	HF88A04
207	EST00170	M62111	HHC199	342	EST00404	M62285	HHC599	398	EST01857	M85349	HF88A03
208	EST00171	M62112	HHC199	343	EST00405	M62286	HHC599	399	EST01858	M85350	HF88A02
209	EST00172	M62113	HHC199	344	EST00406	M62287	HHC599	400	EST01859	M85351	HF88A01
210	EST00173	M62114	HHC199	345	EST00407	M62288	HHC599	401	EST01860	M85352	HF88A00
211	EST00174	M62115	HHC199	346	EST00408	M62289	HHC599	402	EST01861	M85353	HF88A00
212	EST00175	M62116	HHC199	347	EST00409	M62290	HHC599	403	EST01862	M85354	HF88A00
213	EST00176	M62117	HHC199	348	EST00410	M62291	HHC599	404	EST01863	M85355	HF88A00
214	EST00177	M62118	HHC199	349	EST00411	M62292	HHC599	405	EST01864	M85356	HF88A00
215	EST00178	M62119	HHC199	350	EST00412	M62293	HHC599	406	EST01865	M85357	HF88A00
216	EST00179	M62120	HHC199	351	EST00413	M62294	HHC599	407	EST01866	M85358	HF88A00
217	EST00180	M62121	HHC199	352	EST00414	M62295	HHC599	408	EST01867	M85359	HF88A00
218	EST00181	M62122	HHC199	353	EST00415	M62296	HHC599	409	EST01868	M85360	HF88A00
219	EST00182	M62123	HHC199	354	EST00416	M62297	HHC599	410	EST01869	M85361	HF88A00
220	EST00183	M62124	HHC199	355	EST00417	M62298	HHC599	411	EST01870	M85362	HF88A00
221	EST00184	M62125	HHC199	356	EST00418	M62299	HHC599	412	EST01871	M85363	HF88A00
222	EST00185	M62126	HHC199	357	EST00419	M62300	HHC599	413	EST01872	M85364	HF88A00
223	EST00186	M62127	HHC199	358	EST00420	M62301	HHC599	414	EST01873	M85365	HF88A00
224	EST00187	M62128	HHC199	359	EST00421	M62302	HHC599	415	EST01874	M85366	HF88A00
225	EST00188	M62129	HHC199	360	EST00422	M62303	HHC599	416	EST01875	M85367	HF88A00
226	EST00189	M62130	HHC199	361	EST00423	M62304	HHC599	417	EST01876	M85368	HF88A00
227	EST00190	M62131	HHC199	362	EST00424	M62305	HHC599	418	EST01877	M85369	HF88A00
228	EST00191	M62132	HHC199	363	EST00425	M62306	HHC599	419	EST01878	M85370	HF88A00
229	EST00192	M62133	HHC199	364	EST00426	M62307	HHC599	420	EST01879	M85371	HF88A00
230	EST00193	M62134	HHC199	365	EST00427	M62308	HHC599	421	EST01880	M85372	HF88A00
231	EST00194	M62135	HHC199	366	EST00428	M62309	HHC599	422	EST01881	M85373	HF88A00
232	EST00195	M62136	HHC199	367	EST00429	M62310	HHC599	423	EST01882	M85374	HF88A00
233	EST00196	M62137	HHC199	368	EST00430	M62311	HHC599	424	EST01883	M85375	HF88A00
234	EST00197	M62138	HHC199	369	EST00431	M62312	HHC599	425	EST01884	M85376	HF88A00
235	EST00198	M62139	HHC199	370	EST00432	M62313	HHC599	426	EST01885	M85377	HF88A00
236	EST00199	M62140	HHC199	371	EST00433	M62314	HHC599	427	EST01886	M85378	HF88A00
237	EST00200	M62141	HHC199	372	EST00434	M62315	HHC599	428	EST01887	M85379	HF88A00
238	EST00201	M62142	HHC199	373	EST00435	M62316	HHC599	429	EST01888	M85380	HF88A00
239	EST00202	M62143	HHC199	374	EST00436	M62317	HHC599	430	EST01889	M85381	HF88A00
240	EST00203	M62144	HHC199	375	EST00437	M62318	HHC599	431	EST01890	M85382	HF88A00
241	EST00204	M62145	HHC199	376	EST00438	M62319	HHC599	432	EST01891	M85383	HF88A00
242	EST00205	M62146	HHC199	377	EST00439	M62320	HHC599	433	EST01892	M85384	HF88A00
243	EST00206	M62147	HHC199	378	EST00440	M62321	HHC599	434	EST01893	M85385	HF88A00
244	EST00207	M62148	HHC199	379	EST00441	M62322	HHC599	435	EST01894	M85386	HF88A00
245	EST00208	M62149	HHC199	380	EST00442	M62323	HHC599	436	EST01895	M85387	HF88A00
246	EST00209	M62150	HHC199	381	EST00443	M62324	HHC599	437	EST01896	M85388	HF88A00
247	EST00210	M62151	HHC199	382	EST00444	M62325	HHC599	438	EST01897	M85389	HF88A00
248	EST00211	M62152	HHC199	383	EST00445	M62326	HHC599	439	EST01898	M85390	HF88A00
249	EST00212	M62153	HHC199	384	EST00446	M62327	HHC599	440	EST01899	M85391	HF88A00
250	EST00213	M62154	HHC199	385	EST00447	M62328	HHC599	441	EST01900	M85392	HF88A00
251	EST00214	M62155	HHC199	386	EST00448	M62329	HHC599	442	EST01901	M85393	HF88A00
252	EST00215	M62156	HHC199	387	EST00449	M62330	HHC599	443	EST01902	M85394	HF88A00
253	EST00216	M62157	HHC199	388	EST00450	M62331	HHC599	444	EST01903	M85395	HF88A00
254	EST00217	M62158	HHC199	389	EST00451	M62332	HHC599	445	EST01904	M85396	HF88A00
255	EST00218	M62159	HHC199	390	EST00452	M62333	HHC599	446	EST01905	M85397	HF88A00
256	EST00219	M62160	HHC199	391	EST00453	M62334	HHC599	447	EST01906	M85398	HF88A00
257	EST00220	M62161	HHC199	392	EST00454	M62335	HHC599	448	EST01907	M85399	HF88A00
258	EST00221	M62162	HHC199	393	EST00455	M62336	HHC599	449	EST01908	M85400	HF88A00
259	EST00222	M62163	HHC199	394	EST00456	M62337	HHC599	450	EST01909	M85401	HF88A00
260	EST00223	M62164	HHC199	395	EST00457	M62338	HHC599	451	EST01910	M85402	HF88A00
261	EST00224	M62165	HHC199	396	EST00458	M62339	HHC599	452	EST01911	M85403	HF88A00
262	EST00225	M62166	HHC199	397	EST00459	M62340	HHC599	453	EST01912	M85404	HF88A00
263	EST00226	M62167	HHC199	398	EST00460	M62341	HHC599	454	EST01913	M85405	HF88A00
264	EST00227	M62168	HHC199	399	EST00461	M62342	HHC599	455	EST01914	M85406	HF88A00
265	EST00228	M62169	HHC199	400	EST00462	M62343	HHC599	456	EST01915	M85407	HF88A00
266	EST00229	M62170	HHC199	401	EST00463	M62344	HHC599	457	EST01916	M85408	HF88A00
267	EST00230	M62171	HHC199	402	EST00464	M62345	HHC599	458	EST01917	M85409	HF88A00
268	EST00231	M62172	HHC199	403	EST00465	M62346	HHC599	459	EST01918	M85410	HF88A00
269	EST00232	M62173	HHC199	404	EST00466	M62347	HHC599	460	EST01919	M85411	HF88A00
270	EST00233	M62174	HHC199	405	EST00467	M62348	HHC599	461	EST01920	M85412	HF88A00
271	EST00234	M62175	HHC199	406	EST00468	M62349	HHC599	462	EST01921	M85413	HF88A00
272	EST00235	M62176	HHC199	407	EST00469	M62350	HHC599	463	EST01922	M85414	HF88A00
273	EST00236	M62177	HHC199	408	EST00470	M62351	HHC599	464	EST01923	M85415	HF88A00
274	EST00237	M62178	HHC199	409	EST00471	M62352	HHC599	465	EST01924	M85416	HF88A00
275	EST00238	M62179	HHC199	410	EST00472	M62353	HHC599	466	EST01925	M85417	HF88A00
276	EST00239	M62180	HHC199	411	EST00473	M62354	HHC599	467	EST01926	M85418	HF88A00
277	EST00240	M62181	HHC199	412	EST00474	M62355	HHC599	468	EST01927	M85419	HF88A00
278	EST00241	M62182	HHC199	413	EST00475	M62356	HHC599	469	EST01928	M85420	HF88A00
279	EST00242	M62183	HHC199	414	EST00476	M62357	HHC599	470	EST01929	M85421	HF88A00
280	EST00243	M62184	HHC199	415	EST00477	M62358	HHC599	471	EST01930	M85422	HF88A00
281	EST00244	M62185	HHC199	416							

SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone
374	EST00428	W78280	HFBA30	440	EST01454	W7870	HFBA22	507	EST00530	W78382	HFBC81	562	EST00572	W78425	HFBC89	563	EST00573	W78425	HFBC89
375	EST00429	W78281	HFBA32	441	EST00481	W78333	HFBA26	508	EST00531	W78383	HFBC82	563	EST00573	W78425	HFBC89	564	EST00574	W78425	HFBC89
376	EST01436	W78282	HFBA33	442	EST01456	W7872	HFBA27	509	EST00532	W78384	HFBC83	564	EST00574	W78425	HFBC89	565	EST00575	W78425	HFBC89
377	EST00430	W78283	HFBA34	443	EST00482	W78334	HFBA28	510	EST00533	W78385	HFBC84	565	EST00575	W78425	HFBC89	566	EST00576	W78425	HFBC89
378	EST00431	W78284	HFBA35	444	EST00483	W78335	HFBA29	511	EST00534	W78386	HFBC85	566	EST00576	W78425	HFBC89	567	EST00577	W78425	HFBC89
379	EST00432	W78285	HFBA36	445	EST00484	W78336	HFBA30	512	EST00535	W78387	HFBC86	567	EST00577	W78425	HFBC89	568	EST00578	W78425	HFBC89
380	EST01437	W78286	HFBA37	446	EST00485	W78337	HFBA31	513	EST00536	W78388	HFBC87	568	EST00578	W78425	HFBC89	569	EST00579	W78425	HFBC89
381	EST00433	W78287	HFBA38	447	EST00486	W78338	HFBA32	514	EST00537	W78389	HFBC88	569	EST00579	W78425	HFBC89	570	EST00580	W78425	HFBC89
382	EST00434	W78288	HFBA39	448	EST00487	W78339	HFBA33	515	EST00538	W78390	HFBC89	570	EST00580	W78425	HFBC89	571	EST00581	W78425	HFBC89
383	EST00435	W78289	HFBA40	449	EST00488	W78340	HFBA34	516	EST00539	W78391	HFBC90	571	EST00581	W78425	HFBC89	572	EST00582	W78425	HFBC89
384	EST01440	W78290	HFBA41	450	EST00489	W78341	HFBA35	517	EST00540	W78392	HFBC91	572	EST00582	W78425	HFBC89	573	EST00583	W78425	HFBC89
385	EST00436	W78291	HFBA42	451	EST00490	W78342	HFBA36	518	EST00541	W78393	HFBC92	573	EST00583	W78425	HFBC89	574	EST00584	W78425	HFBC89
386	EST00437	W78292	HFBA43	452	EST00491	W78343	HFBA37	519	EST00542	W78394	HFBC93	574	EST00584	W78425	HFBC89	575	EST00585	W78425	HFBC89
387	EST00438	W78293	HFBA44	453	EST00492	W78344	HFBA38	520	EST00543	W78395	HFBC94	575	EST00585	W78425	HFBC89	576	EST00586	W78425	HFBC89
388	EST00439	W78294	HFBA45	454	EST00493	W78345	HFBA39	521	EST00544	W78396	HFBC95	576	EST00586	W78425	HFBC89	577	EST00587	W78425	HFBC89
389	EST01442	W78295	HFBA46	455	EST00494	W78346	HFBA40	522	EST00545	W78397	HFBC96	577	EST00587	W78425	HFBC89	578	EST00588	W78425	HFBC89
390	EST00440	W78296	HFBA47	456	EST00495	W78347	HFBA41	523	EST00546	W78398	HFBC97	578	EST00588	W78425	HFBC89	579	EST00589	W78425	HFBC89
391	EST00441	W78297	HFBA48	457	EST00496	W78348	HFBA42	524	EST00547	W78399	HFBC98	579	EST00589	W78425	HFBC89	580	EST00590	W78425	HFBC89
392	EST00442	W78298	HFBA49	458	EST00497	W78349	HFBA43	525	EST00548	W78400	HFBC99	580	EST00590	W78425	HFBC89	581	EST00591	W78425	HFBC89
393	EST00443	W78299	HFBA50	459	EST00498	W78350	HFBA44	526	EST00549	W78401	HFBC00	581	EST00591	W78425	HFBC89	582	EST00592	W78425	HFBC89
394	EST00444	W78300	HFBA51	460	EST00499	W78351	HFBA45	527	EST00550	W78402	HFBC01	582	EST00592	W78425	HFBC89	583	EST00593	W78425	HFBC89
395	EST01445	W78301	HFBA52	461	EST00500	W78352	HFBA46	528	EST00551	W78403	HFBC02	583	EST00593	W78425	HFBC89	584	EST00594	W78425	HFBC89
396	EST00445	W78302	HFBA53	462	EST00501	W78353	HFBA47	529	EST00552	W78404	HFBC03	584	EST00594	W78425	HFBC89	585	EST00595	W78425	HFBC89
397	EST00446	W78303	HFBA54	463	EST00502	W78354	HFBA48	530	EST00553	W78405	HFBC04	585	EST00595	W78425	HFBC89	586	EST00596	W78425	HFBC89
398	EST00447	W78304	HFBA55	464	EST00503	W78355	HFBA49	531	EST00554	W78406	HFBC05	586	EST00596	W78425	HFBC89	587	EST00597	W78425	HFBC89
399	EST00448	W78305	HFBA56	465	EST00504	W78356	HFBA50	532	EST00555	W78407	HFBC06	587	EST00597	W78425	HFBC89	588	EST00598	W78425	HFBC89
400	EST00449	W78306	HFBA57	466	EST00505	W78357	HFBA51	533	EST00556	W78408	HFBC07	588	EST00598	W78425	HFBC89	589	EST00599	W78425	HFBC89
401	EST00450	W78307	HFBA58	467	EST00506	W78358	HFBA52	534	EST00557	W78409	HFBC08	589	EST00599	W78425	HFBC89	590	EST00600	W78425	HFBC89
402	EST00451	W78308	HFBA59	468	EST00507	W78359	HFBA53	535	EST00558	W78410	HFBC09	590	EST00600	W78425	HFBC89	591	EST00601	W78425	HFBC89
403	EST00452	W78309	HFBA60	469	EST00508	W78360	HFBA54	536	EST00559	W78411	HFBC10	591	EST00601	W78425	HFBC89	592	EST00602	W78425	HFBC89
404	EST00453	W78310	HFBA61	470	EST00509	W78361	HFBA55	537	EST00560	W78412	HFBC11	592	EST00602	W78425	HFBC89	593	EST00603	W78425	HFBC89
405	EST00454	W78311	HFBA62	471	EST00510	W78362	HFBA56	538	EST00561	W78413	HFBC12	593	EST00603	W78425	HFBC89	594	EST00604	W78425	HFBC89
406	EST00455	W78312	HFBA63	472	EST00511	W78363	HFBA57	539	EST00562	W78414	HFBC13	594	EST00604	W78425	HFBC89	595	EST00605	W78425	HFBC89
407	EST00456	W78313	HFBA64	473	EST00512	W78364	HFBA58	540	EST00563	W78415	HFBC14	595	EST00605	W78425	HFBC89	596	EST00606	W78425	HFBC89
408	EST00457	W78314	HFBA65	474	EST00513	W78365	HFBA59	541	EST00564	W78416	HFBC15	596	EST00606	W78425	HFBC89	597	EST00607	W78425	HFBC89
409	EST01444	W78315	HFBA66	475	EST00514	W78366	HFBA60	542	EST00565	W78417	HFBC16	597	EST00607	W78425	HFBC89	598	EST00608	W78425	HFBC89
410	EST00458	W78316	HFBA67	476	EST00515	W78367	HFBA61	543	EST00566	W78418	HFBC17	598	EST00608	W78425	HFBC89	599	EST00609	W78425	HFBC89
411	EST00459	W78317	HFBA68	477	EST00516	W78368	HFBA62	544	EST00567	W78419	HFBC18	599	EST00609	W78425	HFBC89	600	EST00610	W78425	HFBC89
412	EST00460	W78318	HFBA69	478	EST00517	W78369	HFBA63	545	EST00568	W78420	HFBC19	600	EST00610	W78425	HFBC89	601	EST00611	W78425	HFBC89
413	EST00461	W78319	HFBA70	479	EST00518	W78370	HFBA64	546	EST00569	W78421	HFBC20	601	EST00611	W78425	HFBC89	602	EST00612	W78425	HFBC89
414	EST00462	W78320	HFBA71	480	EST00519	W78371	HFBA65	547	EST00570	W78422	HFBC21	602	EST00612	W78425	HFBC89	603	EST00613	W78425	HFBC89
415	EST00463	W78321	HFBA72	481	EST00520	W78372	HFBA66	548	EST00571	W78423	HFBC22	603	EST00613	W78425	HFBC89	604	EST00614	W78425	HFBC89
416	EST00464	W78322	HFBA73	482	EST00521	W78373	HFBA67	549	EST00572	W78424	HFBC23	604	EST00614	W78425	HFBC89	605	EST00615	W78425	HFBC89
417	EST00465	W78323	HFBA74	483	EST00522	W78374	HFBA68	550	EST00573	W78425	HFBC24	605	EST00615	W78425	HFBC89	606	EST00616	W78425	HFBC89
418	EST00466	W78324	HFBA75	484	EST00523	W78375	HFBA69	551	EST00574	W78426	HFBC25	606	EST00616	W78425	HFBC89	607	EST00617	W78425	HFBC89
419	EST00467	W78325	HFBA76	485	EST00524	W78376	HFBA70	552	EST00575	W78427	HFBC26	607	EST00617	W78425	HFBC89	608	EST00618	W78425	HFBC89
420	EST00468	W78326	HFBA77	486	EST00525	W78377	HFBA71	553	EST00576	W78428	HFBC27	608	EST00618	W78425	HFBC89	609	EST00619	W78425	HFBC89
421	EST00469	W78327	HFBA78	487	EST00526	W78378	HFBA72	554	EST00577	W78429	HFBC28	609	EST00619	W78425	HFBC89	610	EST00620	W78425	HFBC89
422	EST00470	W78328	HFBA79	488	EST00527	W78379	HFBA73	555	EST00578	W78430	HFBC29	610	EST00620	W78425	HFBC89	611	EST00621	W78425	HFBC89
423	EST00471	W78329	HFBA80	489	EST00528	W78380	HFBA74	556	EST00579	W78431	HFBC30	611	EST00621	W78425	HFBC89	612	EST00622	W78425	HFBC89
424	EST00472	W78330	HFBA81	490	EST00529	W78381	HFBA75	557	EST00580	W78432	HFBC31	612	EST00622	W78425	HFBC89	613	EST00623	W78425	HFBC89
425	EST00473	W78331	HFBA82	491	EST00530	W78382	HFBA76	558	EST00581	W78433	HFBC32	613	EST00623	W78425	HFBC89	614	EST00624	W78425	HFBC89
426	EST00474	W78332	HFBA83	492	EST00531	W78383	HFBA77	559	EST00582	W78434	HFBC33	614	EST00624	W78425	HFBC89	615	EST00625	W78425	HFBC89
427	EST00475	W78333	HFBA84	493	EST00532	W78384	HFBA78	560	EST00583	W78435	HFBC34	615	EST00625	W78425	HFBC89	616	EST00626	W78425	HFBC89
428	EST00476	W78334	HFBA85	494	EST00533	W78385	HFBA79	561	EST00584	W78436	HFBC35	616	EST00626	W78425	HFBC89	617	EST00627	W78425	HFBC89
429	EST00477	W78335	HFBA86	495	EST00534	W78386	HFBA80	562	EST00585	W78437	HFBC36	617	EST00627	W78425	HFBC89	618	EST00628	W78425	HFBC89
430	EST00478	W78336	HFBA87	496	EST00535	W78387	HFBA81	563	EST00586	W78438	HFBC37	618	EST00628	W78425	HFBC89	619	EST00629	W78425	HFBC89
431	EST00479	W78337	HFBA88	497	EST00536	W78388	HFBA82	564	EST00587	W78439	HFBC38	619	EST00629	W78425	HFBC89	620	EST00630	W78425	HFBC89
432	EST00480	W78338	HFBA89	498	EST00537	W78389	HFBA83	565	EST00588	W78440	HFBC39	620	EST00630	W78425	HFBC89	621	EST00631	W78425	HFBC89
433	EST00481	W78339	HFBA90	499	EST00538	W78390	HFBA84	566	EST00589	W78441	HFBC40	621	EST00631	W78425	HFBC89	622	EST00632	W78425	HFBC89
434	EST00482	W78340	HFBA9																

564	EST000574	W78426	HFBCB92	Clone	EST01521	W78337	HFBCD87	Clone
565	EST000575	W78427	HFBCB93		EST000680	W78338	HFBCD88	
566	EST000576	W78428	HFBCB94		EST000681	W78339	HFBCD89	
567	EST000577	W78429	HFBCB95		EST000682	W78340	HFBCD90	
568	EST000578	W78430	HFBCB96		EST000683	W78341	HFBCD91	
569	EST000579	W78431	HFBCB97		EST000684	W78342	HFBCD92	
570	EST000580	W78432	HFBCB98		EST000685	W78343	HFBCD93	
571	EST000581	W78433	HFBCB99		EST000686	W78344	HFBCD94	
572	EST000582	W78434	HFBCB00		EST000687	W78345	HFBCD95	
573	EST000583	W78435	HFBCB01		EST000688	W78346	HFBCD96	
574	EST000584	W78436	HFBCB02		EST000689	W78347	HFBCD97	
575	EST000585	W78437	HFBCB03		EST000690	W78348	HFBCD98	
576	EST000586	W78438	HFBCB04		EST000691	W78349	HFBCD99	
577	EST000587	W78439	HFBCB05		EST000692	W78350	HFBCD00	
578	EST000588	W78440	HFBCB06		EST000693	W78351	HFBCD01	
579	EST000589	W78441	HFBCB07		EST000694	W78352	HFBCD02	
580	EST000590	W78442	HFBCB08		EST000695	W78353	HFBCD03	
581	EST000591	W78443	HFBCB09		EST000696	W78354	HFBCD04	
582	EST000592	W78444	HFBCB10		EST000697	W78355	HFBCD05	
583	EST000593	W78445	HFBCB11		EST000698	W78356	HFBCD06	
584	EST000594	W78446	HFBCB12		EST000699	W78357	HFBCD07	
585	EST000595	W78447	HFBCB13		EST000700	W78358	HFBCD08	
586	EST000596	W78448	HFBCB14		EST000701	W78359	HFBCD09	
587	EST000597	W78449	HFBCB15		EST000702	W78360	HFBCD10	
588	EST000598	W78450	HFBCB16		EST000703	W78361	HFBCD11	
589	EST000599	W78451	HFBCB17		EST000704	W78362	HFBCD12	
590	EST000600	W78452	HFBCB18		EST000705	W78363	HFBCD13	
591	EST000601	W78453	HFBCB19		EST000706	W78364	HFBCD14	
592	EST000602	W78454	HFBCB20		EST000707	W78365	HFBCD15	
593	EST000603	W78455	HFBCB21		EST000708	W78366	HFBCD16	
594	EST000604	W78456	HFBCB22		EST000709	W78367	HFBCD17	
595	EST000605	W78457	HFBCB23		EST000710	W78368	HFBCD18	
596	EST000606	W78458	HFBCB24		EST000711	W78369	HFBCD19	
597	EST000607	W78459	HFBCB25		EST000712	W78370	HFBCD20	
598	EST000608	W78460	HFBCB26		EST000713	W78371	HFBCD21	
599	EST000609	W78461	HFBCB27		EST000714	W78372	HFBCD22	
600	EST000610	W78462	HFBCB28		EST000715	W78373	HFBCD23	
601	EST000611	W78463	HFBCB29		EST000716	W78374	HFBCD24	
602	EST000612	W78464	HFBCB30		EST000717	W78375	HFBCD25	
603	EST000613	W78465	HFBCB31		EST000718	W78376	HFBCD26	
604	EST000614	W78466	HFBCB32		EST000719	W78377	HFBCD27	
605	EST000615	W78467	HFBCB33		EST000720	W78378	HFBCD28	
606	EST000616	W78468	HFBCB34		EST000721	W78379	HFBCD29	
607	EST000617	W78469	HFBCB35		EST000722	W78380	HFBCD30	
608	EST000618	W78470	HFBCB36		EST000723	W78381	HFBCD31	
609	EST000619	W78471	HFBCB37		EST000724	W78382	HFBCD32	
610	EST000620	W78472	HFBCB38		EST000725	W78383	HFBCD33	
611	EST000621	W78473	HFBCB39		EST000726	W78384	HFBCD34	
612	EST000622	W78474	HFBCB40		EST000727	W78385	HFBCD35	
613	EST000623	W78475	HFBCB41		EST000728	W78386	HFBCD36	
614	EST000624	W78476	HFBCB42		EST000729	W78387	HFBCD37	
615	EST000625	W78477	HFBCB43		EST000730	W78388	HFBCD38	
616	EST000626	W78478	HFBCB44		EST000731	W78389	HFBCD39	
617	EST000627	W78479	HFBCB45		EST000732	W78390	HFBCD40	
618	EST000628	W78480	HFBCB46		EST000733	W78391	HFBCD41	
619	EST000629	W78481	HFBCB47		EST000734	W78392	HFBCD42	
620	EST000630	W78482	HFBCB48		EST000735	W78393	HFBCD43	
621	EST000631	W78483	HFBCB49		EST000736	W78394	HFBCD44	
622	EST000632	W78484	HFBCB50		EST000737	W78395	HFBCD45	
623	EST000633	W78485	HFBCB51		EST000738	W78396	HFBCD46	
624	EST000634	W78486	HFBCB52		EST000739	W78397	HFBCD47	
625	EST000635	W78487	HFBCB53		EST000740	W78398	HFBCD48	
626	EST000636	W78488	HFBCB54		EST000741	W78399	HFBCD49	
627	EST000637	W78489	HFBCB55		EST000742	W78400	HFBCD50	
628	EST000638	W78490	HFBCB56		EST000743	W78401	HFBCD51	
629	EST000639	W78491	HFBCB57		EST000744	W78402	HFBCD52	

SUBSTITUTE SHEET

-101-

SEQ ID	EST#	CS#	Clone	SEQ ID	EST#	CS#	Clone	SEQ ID	EST#	CS#	Clone	SEQ ID	EST#	CS#	Clone
753	EST00721	W78573	HFBC66	885	EST01885	M85371	HFBC112	941	EST01941	M85428	HFBC108	941	EST01941	M85428	HFBC108
754	EST00722	W78574	HFBC67	886	EST01886	M85372	HFBC113	942	EST01942	M85429	HFBC109	942	EST01942	M85429	HFBC109
755	EST00723	W78575	HFBC68	887	EST01887	M85373	HFBC114	943	EST01943	M85430	HFBC110	943	EST01943	M85430	HFBC110
756	EST01541	W77957	HFBC69	888	EST01888	M85374	HFBC115	944	EST01944	M85431	HFBC111	944	EST01944	M85431	HFBC111
757	EST00724	W77958	HFBC70	889	EST01889	M85375	HFBC116	945	EST01945	M85432	HFBC112	945	EST01945	M85432	HFBC112
758	EST00725	W78576	HFBC71	890	EST01890	M85376	HFBC117	946	EST01946	M85433	HFBC113	946	EST01946	M85433	HFBC113
759	EST00726	W78577	HFBC72	891	EST01891	M85377	HFBC118	947	EST01947	M85434	HFBC114	947	EST01947	M85434	HFBC114
760	EST01544	W77960	HFBC73	892	EST01892	M85378	HFBC119	948	EST01948	M85435	HFBC115	948	EST01948	M85435	HFBC115
761	EST00727	W78578	HFBC74	893	EST01893	M85379	HFBC120	949	EST01949	M85436	HFBC116	949	EST01949	M85436	HFBC116
762	EST00728	W78579	HFBC75	894	EST01894	M85380	HFBC121	950	EST01950	M85437	HFBC117	950	EST01950	M85437	HFBC117
763	EST00729	W78580	HFBC76	895	EST01895	M85381	HFBC122	951	EST01951	M85438	HFBC118	951	EST01951	M85438	HFBC118
764	EST00730	W78581	HFBC77	896	EST01896	M85382	HFBC123	952	EST01952	M85439	HFBC119	952	EST01952	M85439	HFBC119
765	EST00731	W78582	HFBC78	897	EST01897	M85383	HFBC124	953	EST01953	M85440	HFBC120	953	EST01953	M85440	HFBC120
766	EST00732	W78583	HFBC79	898	EST01898	M85384	HFBC125	954	EST01954	M85441	HFBC121	954	EST01954	M85441	HFBC121
767	EST00733	W78584	HFBC80	899	EST01899	M85385	HFBC126	955	EST01955	M85442	HFBC122	955	EST01955	M85442	HFBC122
768	EST00734	W78585	HFBC81	900	EST01900	M85386	HFBC127	956	EST01956	M85443	HFBC123	956	EST01956	M85443	HFBC123
769	EST00735	W78586	HFBC82	901	EST01901	M85387	HFBC128	957	EST01957	M85444	HFBC124	957	EST01957	M85444	HFBC124
770	EST01546	W77962	HFBC83	902	EST01902	M85388	HFBC129	958	EST01958	M85445	HFBC125	958	EST01958	M85445	HFBC125
771	EST00736	W78587	HFBC84	903	EST01903	M85389	HFBC130	959	EST01959	M85446	HFBC126	959	EST01959	M85446	HFBC126
772	EST00737	W78588	HFBC85	904	EST01904	M85390	HFBC131	960	EST01960	M85447	HFBC127	960	EST01960	M85447	HFBC127
773	EST01547	W77963	HFBC86	905	EST01905	M85391	HFBC132	961	EST01961	M85448	HFBC128	961	EST01961	M85448	HFBC128
774	EST00738	W78589	HFBC87	906	EST01906	M85392	HFBC133	962	EST01962	M85449	HFBC129	962	EST01962	M85449	HFBC129
775	EST00739	W78590	HFBC88	907	EST01907	M85393	HFBC134	963	EST01963	M85450	HFBC130	963	EST01963	M85450	HFBC130
776	EST00740	W78591	HFBC89	908	EST01908	M85394	HFBC135	964	EST01964	M85451	HFBC131	964	EST01964	M85451	HFBC131
777	EST00741	W78592	HFBC90	909	EST01909	M85395	HFBC136	965	EST01965	M85452	HFBC132	965	EST01965	M85452	HFBC132
778	EST00742	W78593	HFBC91	910	EST01910	M85396	HFBC137	966	EST01966	M85453	HFBC133	966	EST01966	M85453	HFBC133
779	EST00743	W78594	HFBC92	911	EST01911	M85397	HFBC138	967	EST01967	M85454	HFBC134	967	EST01967	M85454	HFBC134
780	EST01549	W77964	HFBC93	912	EST01912	M85398	HFBC139	968	EST01968	M85455	HFBC135	968	EST01968	M85455	HFBC135
781	EST01550	W77965	HFBC94	913	EST01913	M85399	HFBC140	969	EST01969	M85456	HFBC136	969	EST01969	M85456	HFBC136
782	EST01551	W77966	HFBC95	914	EST01914	M85400	HFBC141	970	EST01970	M85457	HFBC137	970	EST01970	M85457	HFBC137
783	EST01552	W77967	HFBC96	915	EST01915	M85401	HFBC142	971	EST01971	M85458	HFBC138	971	EST01971	M85458	HFBC138
784	EST01553	W77968	HFBC97	916	EST01916	M85402	HFBC143	972	EST01972	M85459	HFBC139	972	EST01972	M85459	HFBC139
785	EST00744	W78595	HFBC98	917	EST01917	M85403	HFBC144	973	EST01973	M85460	HFBC140	973	EST01973	M85460	HFBC140
786	EST00745	W78596	HFBC99	918	EST01918	M85404	HFBC145	974	EST01974	M85461	HFBC141	974	EST01974	M85461	HFBC141
787	EST00746	W78597	HFBC00	919	EST01919	M85405	HFBC146	975	EST01975	M85462	HFBC142	975	EST01975	M85462	HFBC142
788	EST00747	W78598	HFBC01	920	EST01920	M85406	HFBC147	976	EST01976	M85463	HFBC143	976	EST01976	M85463	HFBC143
789	EST00748	W78599	HFBC02	921	EST01921	M85407	HFBC148	977	EST01977	M85464	HFBC144	977	EST01977	M85464	HFBC144
790	EST01554	W77970	HFBC03	922	EST01922	M85408	HFBC149	978	EST01978	M85465	HFBC145	978	EST01978	M85465	HFBC145
791	EST00749	W78600	HFBC04	923	EST01923	M85409	HFBC150	979	EST01979	M85466	HFBC146	979	EST01979	M85466	HFBC146
792	EST00750	W78601	HFBC05	924	EST01924	M85410	HFBC151	980	EST01980	M85467	HFBC147	980	EST01980	M85467	HFBC147
793	EST01555	W77971	HFBC06	925	EST01925	M85411	HFBC152	981	EST01981	M85468	HFBC148	981	EST01981	M85468	HFBC148
794	EST00751	W78602	HFBC07	926	EST01926	M85412	HFBC153	982	EST01982	M85469	HFBC149	982	EST01982	M85469	HFBC149
795	EST00752	W78603	HFBC08	927	EST01927	M85413	HFBC154	983	EST01983	M85470	HFBC150	983	EST01983	M85470	HFBC150
796	EST00753	W78604	HFBC09	928	EST01928	M85414	HFBC155	984	EST01984	M85471	HFBC151	984	EST01984	M85471	HFBC151
797	EST00754	W78605	HFBC10	929	EST01929	M85415	HFBC156	985	EST01985	M85472	HFBC152	985	EST01985	M85472	HFBC152
798	EST00755	W78606	HFBC11	930	EST01930	M85416	HFBC157	986	EST01986	M85473	HFBC153	986	EST01986	M85473	HFBC153
799	EST00756	W78607	HFBC12	931	EST01931	M85417	HFBC158	987	EST01987	M85474	HFBC154	987	EST01987	M85474	HFBC154
800	EST00757	W78608	HFBC13	932	EST01932	M85418	HFBC159	988	EST01988	M85475	HFBC155	988	EST01988	M85475	HFBC155
801	EST00758	W78609	HFBC14	933	EST01933	M85419	HFBC160	989	EST01989	M85476	HFBC156	989	EST01989	M85476	HFBC156
802	EST00759	W78610	HFBC15	934	EST01934	M85420	HFBC161	990	EST01990	M85477	HFBC157	990	EST01990	M85477	HFBC157
803	EST00760	W78611	HFBC16	935	EST01935	M85421	HFBC162	991	EST01991	M85478	HFBC158	991	EST01991	M85478	HFBC158
804	EST00761	W78612	HFBC17	936	EST01936	M85422	HFBC163	992	EST01992	M85479	HFBC159	992	EST01992	M85479	HFBC159
805	EST00762	W78613	HFBC18	937	EST01937	M85423	HFBC164	993	EST01993	M85480	HFBC160	993	EST01993	M85480	HFBC160
806	EST00763	W78614	HFBC19	938	EST01938	M85424	HFBC165	994	EST01994	M85481	HFBC161	994	EST01994	M85481	HFBC161
807	EST00764	W78615	HFBC20	939	EST01939	M85425	HFBC166	995	EST01995	M85482	HFBC162	995	EST01995	M85482	HFBC162
808	EST01854	W78616	HFBC21	940	EST01940	M85426	HFBC201	996	EST01996	M85483	HFBC163	996	EST01996	M85483	HFBC163
809	EST00765	W78617	HFBC22	941	EST01941	M85427	HFBC202	997	EST01997	M85484	HFBC164	997	EST01997	M85484	HFBC164
810	EST00766	W78618	HFBC23	942	EST01942	M85428	HFBC203	998	EST01998	M85485	HFBC165	998	EST01998	M85485	HFBC165
811	EST01855	W78619	HFBC24	943	EST01943	M85429	HFBC204	999	EST01999	M85486	HFBC166	999	EST01999	M85486	HFBC166
812	EST00767	W78620	HFBC25	944	EST01944	M85430	HFBC205	999	EST01999	M85487	HFBC206	999	EST01999	M85487	HFBC206
813	EST00768	W78621	HFBC26	945	EST01945	M85431	HFBC207	999	EST01999	M85488	HFBC208	999	EST01999	M85488	HFBC208
814	EST00769	W78622	HFBC27	946	EST01946	M85432	HFBC209	999	EST01999	M85489	HFBC210	999	EST01999	M85489	HFBC210
815	EST00770	W78623	HFBC28	947	EST01947	M85433	HFBC211	999	EST01999	M85490	HFBC212	999	EST01999	M85490	HFBC212
816	EST01856	W78624	HFBC29	948	EST01948	M85434	HFBC213	999	EST01999	M85491	HFBC214	999	EST01999	M85491	HFBC214
817	EST01857	W78625	HFBC30	949	EST01949	M85435	HFBC215	999	EST01999	M85492	HFBC216	999	EST01999	M85492	HFBC216
818	EST01858	W78626	HFBC31	950	EST01950	M85436	HFBC217	999	EST01999	M85493	HFBC218	999	EST01999	M85493	HFBC218

SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone
942	EST01947	M85431	HFBC111	1074	EST02092	M85576	HFBC129
943	EST01948	M85432	HFBC112	1075	EST02093	M85577	HFBC130
944	EST01949	M85433	HFBC113	1076	EST02094	M85578	HFBC131
945	EST01950	M85434	HFBC114	1077	EST02095	M85580	HFBC136
946	EST01951	M85435	HFBC115	1078	EST02096	M85581	HFBC137
947	EST01952	M85436	HFBC116	1079	EST02097	M85582	HFBC138
948	EST01953	M85437	HFBC117	1080	EST02098	M85583	HFBC139
949	EST01954	M85438	HFBC118	1081	EST02099	M85584	HFBC140
950	EST01955	M85439	HFBC119	1082	EST02100	M85585	HFBC141
951	EST01956	M85440	HFBC120	1083	EST02101	M85586	HFBC142
952	EST01957	M85441	HFBC121	1084	EST02102	M85587	HFBC143
953	EST01958	M85442	HFBC122	1085	EST02103	M85588	HFBC144
954	EST01959	M85443	HFBC123	1086	EST02104	M85589	HFBC145
955	EST01960	M85444	HFBC124	1087	EST02105	M85590	HFBC148
956	EST01961	M85445	HFBC125	1088	EST02106	M85591	HFBC149
957	EST01962	M85446	HFBC126	1089	EST02107	M85592	HFBC154
958	EST01963	M85447	HFBC127	1090	EST02108	M85593	HFBC155
959	EST01964	M85448	HFBC128	1091	EST02109	M85594	HFBC156
960	EST01965	M85449	HFBC129	1092	EST02110	M85595	HFBC157
961	EST01966	M85450	HFBC130	1093	EST02111	M85596	HFBC158
962	EST01967	M85451	HFBC131	1094	EST02112	M85597	HFBC159
963	EST01968	M85452	HFBC132	1095	EST02113	M85598	HFBC160
964	EST01969	M85453	HFBC133	1096	EST02114	M85599	HFBC161
965	EST01970	M85454	HFBC134	1097	EST02115	M85600	HFBC162
966	EST01971	M85455	HFBC135	1098	EST02116	M85601	HFBC163
967	EST01972	M85456	HFBC136	1099	EST02117	M85602	HFBC164
968	EST01973	M85457	HFBC137	1100	EST02118	M85603	HFBC165
969	EST01974	M85458	HFBC138	1101	EST02119	M85604	HFBC166
970	EST01975	M85459	HFBC139	1102	EST02120	M85605	HFBC167
971	EST01976	M85460	HFBC140	1103	EST02121	M85606	HFBC171
972	EST01977	M85461	HFBC141	1104	EST02122	M85607	HFBC172
973	EST01978	M85462	HFBC142	1105	EST02123	M85608	HFBC175
974	EST01979	M85463	HFBC143	1106	EST02124	M85609	HFBC176
975	EST01980	M85464	HFBC144	1107	EST02125	M85610	HFBC177
976	EST01981	M85465	HFBC145	1108	EST02126	M85611	HFBC181
977	EST01982	M85466	HFBC146	1109	EST02127	M85612	HFBC183
978	EST01983	M85467	HFBC147	1110	EST02128	M85613	HFBC184
979	EST01984	M85468	HFBC148	1111	EST02129	M85614	HFBC187
980	EST01985	M85469	HFBC149	1112	EST02130	M85615	HFBC188
981	EST01986	M85470	HFBC150	1113	EST02131	M85616	HFBC189
982	EST01987	M85471	HFBC151	1114	EST02132	M85617	HFBC190
983	EST01988	M85472	HFBC152	1115	EST02133	M85618	HFBC191
984	EST01989	M85473	HFBC153	1116	EST02134	M85619	HFBC192
985	EST01990	M85474	HFBC154	1117	EST02135	M85620	HFBC193
986	EST01991	M85475	HFBC155	1118	EST02136	M85621	HFBC194
987	EST01992	M85476	HFBC156	1119	EST02137	M85622	HFBC195
988	EST01993	M85477	HFBC157	1120	EST02138	M85623	HFBC196
989	EST01994	M85478	HFBC158	1121	EST02139	M85624	HFBC197
990	EST01995	M85479	HFBC159	1122	EST02140	M85625	HFBC198
991	EST01996	M85480	HFBC160	1123	EST02141	M85626	HFBC199
992	EST01997	M85481	HFBC161	1124	EST02142	M85627	HFBC200
993	EST01998	M85482	HFBC162	1125	EST02143	M85628	HFBC201
994	EST01999	M85483	HFBC163	1126	EST02144	M85629	HFBC202
995	EST02000	M85484	HFBC164	1127	EST02145	M85630	HFBC203
996	EST02001	M85485	HFBC165	1128	EST02146	M85631	HFBC204
997	EST02002	M85486	HFBC166	1129	EST02147	M85632	HFBC205
998	EST02003	M85487	HFBC167	1130	EST02148	M85633	HFBC206
999	EST02004	M85488	HFBC168				
1000	EST02005	M85489	HFBC169				
1001	EST02006	M85490	HFBC170				
1002	EST02007	M85491	HFBC171				
1003	EST02008	M85492	HFBC172				
1004	EST02009	M85493	HFBC173				
1005	EST02010	M85494	HFBC174				
1006	EST02011	M85495	HFBC175				
1007	EST02012	M85496	HFBC176				
1008	EST02013	M85497	HFBC177				
1009	EST02014	M85498	HFBC178				
1010	EST02015	M85499	HFBC179				
1011	EST02016	M85500	HFBC180				
1012	EST02017	M85501	HFBC181				
1013	EST02018	M85502	HFBC182				
1014	EST02019	M85503	HFBC183				
1015	EST02020	M85504	HFBC184				
1016	EST02021	M85505	HFBC185				

SUBSTITUTE SHEET

-103-

SEQ ID	EST#	CS#	Clone	SEQ ID	EST#	CS#	Clone	SEQ ID	EST#	CS#	Clone
131	EST02151	M85634	HFBC14	262	EST02281	M85760	HFBC124	302	EST02314	M85803	HFBC190
132	EST02152	M85635	HFBC18	263	EST02282	M85761	HFBC125	303	EST02315	M85804	HFBC191
133	EST02153	M85636	HFBC12	264	EST02283	M85762	HFBC126	304	EST02316	M85805	HFBC192
134	EST02154	M85637	HFBC22	265	EST02284	M85763	HFBC127	305	EST02317	M85806	HFBC193
135	EST02155	M85638	HFBC23	266	EST02285	M85764	HFBC128	306	EST02318	M85807	HFBC194
136	EST02156	M85639	HFBC24	267	EST02286	M85765	HFBC129	307	EST02319	M85808	HFBC195
137	EST02157	M85640	HFBC25	268	EST02287	M85766	HFBC130	308	EST02320	M85809	HFBC196
138	EST02158	M85641	HFBC26	269	EST02288	M85767	HFBC131	309	EST02321	M85810	HFBC197
139	EST02159	M85642	HFBC27	270	EST02289	M85768	HFBC132	310	EST02322	M85811	HFBC198
140	EST02160	M85643	HFBC28	271	EST02290	M85769	HFBC133	311	EST02323	M85812	HFBC199
141	EST02161	M85644	HFBC29	272	EST02291	M85770	HFBC134	312	EST02324	M85813	HFBC200
142	EST02162	M85645	HFBC30	273	EST02292	M85771	HFBC135	313	EST02325	M85814	HFBC201
143	EST02163	M85646	HFBC31	274	EST02293	M85772	HFBC136	314	EST02326	M85815	HFBC202
144	EST02164	M85647	HFBC32	275	EST02294	M85773	HFBC137	315	EST02327	M85816	HFBC203
145	EST02165	M85648	HFBC33	276	EST02295	M85774	HFBC138	316	EST02328	M85817	HFBC204
146	EST02166	M85649	HFBC34	277	EST02296	M85775	HFBC139	317	EST02329	M85818	HFBC205
147	EST02167	M85650	HFBC35	278	EST02297	M85776	HFBC140	318	EST02330	M85819	HFBC206
148	EST02168	M85651	HFBC36	279	EST02298	M85777	HFBC141	319	EST02331	M85820	HFBC207
149	EST02169	M85652	HFBC37	280	EST02299	M85778	HFBC142	320	EST02332	M85821	HFBC208
150	EST02170	M85653	HFBC38	281	EST02300	M85779	HFBC143	321	EST02333	M85822	HFBC209
151	EST02171	M85654	HFBC39	282	EST02301	M85780	HFBC144	322	EST02334	M85823	HFBC210
152	EST02172	M85655	HFBC40	283	EST02302	M85781	HFBC145	323	EST02335	M85824	HFBC211
153	EST02173	M85656	HFBC41	284	EST02303	M85782	HFBC146	324	EST02336	M85825	HFBC212
154	EST02174	M85657	HFBC42	285	EST02304	M85783	HFBC147	325	EST02337	M85826	HFBC213
155	EST02175	M85658	HFBC43	286	EST02305	M85784	HFBC148	326	EST02338	M85827	HFBC214
156	EST02176	M85659	HFBC44	287	EST02306	M85785	HFBC149	327	EST02339	M85828	HFBC215
157	EST02177	M85660	HFBC45	288	EST02307	M85786	HFBC150	328	EST02340	M85829	HFBC216
158	EST02178	M85661	HFBC46	289	EST02308	M85787	HFBC151	329	EST02341	M85830	HFBC217
159	EST02179	M85662	HFBC47	290	EST02309	M85788	HFBC152	330	EST02342	M85831	HFBC218
160	EST02180	M85663	HFBC48	291	EST02310	M85789	HFBC153	331	EST02343	M85832	HFBC219
161	EST02181	M85664	HFBC49	292	EST02311	M85790	HFBC154	332	EST02344	M85833	HFBC220
162	EST02182	M85665	HFBC50	293	EST02312	M85791	HFBC155	333	EST02345	M85834	HFBC221
163	EST02183	M85666	HFBC51	294	EST02313	M85792	HFBC156	334	EST02346	M85835	HFBC222
164	EST02184	M85667	HFBC52	295	EST02314	M85793	HFBC157	335	EST02347	M85836	HFBC223
165	EST02185	M85668	HFBC53	296	EST02315	M85794	HFBC158	336	EST02348	M85837	HFBC224
166	EST02186	M85669	HFBC54	297	EST02316	M85795	HFBC159	337	EST02349	M85838	HFBC225
167	EST02187	M85670	HFBC55	298	EST02317	M85796	HFBC160	338	EST02350	M85839	HFBC226
168	EST02188	M85671	HFBC56	299	EST02318	M85797	HFBC161	339	EST02351	M85840	HFBC227
169	EST02189	M85672	HFBC57	300	EST02319	M85798	HFBC162	340	EST02352	M85841	HFBC228
170	EST02190	M85673	HFBC58	301	EST02320	M85799	HFBC163	341	EST02353	M85842	HFBC229
171	EST02191	M85674	HFBC59	302	EST02321	M85800	HFBC164	342	EST02354	M85843	HFBC230
172	EST02192	M85675	HFBC60	303	EST02322	M85801	HFBC165	343	EST02355	M85844	HFBC231
173	EST02193	M85676	HFBC61	304	EST02323	M85802	HFBC166	344	EST02356	M85845	HFBC232
174	EST02194	M85677	HFBC62	305	EST02324	M85803	HFBC167	345	EST02357	M85846	HFBC233
175	EST02195	M85678	HFBC63	306	EST02325	M85804	HFBC168	346	EST02358	M85847	HFBC234
176	EST02196	M85679	HFBC64	307	EST02326	M85805	HFBC169	347	EST02359	M85848	HFBC235
177	EST02197	M85680	HFBC65	308	EST02327	M85806	HFBC170	348	EST02360	M85849	HFBC236
178	EST02198	M85681	HFBC66	309	EST02328	M85807	HFBC171	349	EST02361	M85850	HFBC237
179	EST02199	M85682	HFBC67	310	EST02329	M85808	HFBC172	350	EST02362	M85851	HFBC238
180	EST02200	M85683	HFBC68	311	EST02330	M85809	HFBC173	351	EST02363	M85852	HFBC239
181	EST02201	M85684	HFBC69	312	EST02331	M85810	HFBC174	352	EST02364	M85853	HFBC240
182	EST02202	M85685	HFBC70	313	EST02332	M85811	HFBC175	353	EST02365	M85854	HFBC241
183	EST02203	M85686	HFBC71	314	EST02333	M85812	HFBC176	354	EST02366	M85855	HFBC242
184	EST02204	M85687	HFBC72	315	EST02334	M85813	HFBC177	355	EST02367	M85856	HFBC243
185	EST02205	M85688	HFBC73	316	EST02335	M85814	HFBC178	356	EST02368	M85857	HFBC244
186	EST02206	M85689	HFBC74	317	EST02336	M85815	HFBC179	357	EST02369	M85858	HFBC245
187	EST02207	M85690	HFBC75	318	EST02337	M85816	HFBC180	358	EST02370	M85859	HFBC246
188	EST02208	M85691	HFBC76	319	EST02338	M85817	HFBC181	359	EST02371	M85860	HFBC247
189	EST02209	M85692	HFBC77	320	EST02339	M85818	HFBC182	360	EST02372	M85861	HFBC248
190	EST02210	M85693	HFBC78	321	EST02340	M85819	HFBC183	361	EST02373	M85862	HFBC249
191	EST02211	M85694	HFBC79	322	EST02341	M85820	HFBC184	362	EST02374	M85863	HFBC250
192	EST02212	M85695	HFBC80	323	EST02342	M85821	HFBC185	363	EST02375	M85864	HFBC251
193	EST02213	M85696	HFBC81	324	EST02343	M85822	HFBC186	364	EST02376	M85865	HFBC252
194	EST02214	M85697	HFBC82	325	EST02344	M85823	HFBC187	365	EST02377	M85866	HFBC253
195	EST02215	M85698	HFBC83	326	EST02345	M85824	HFBC188	366	EST02378	M85867	HFBC254
196	EST02216	M85699	HFBC84	327	EST02346	M85825	HFBC189	367	EST02379	M85868	HFBC255
197	EST02217	M85700	HFBC85	328	EST02347	M85826	HFBC190	368	EST02380	M85869	HFBC256
198	EST02218	M85701	HFBC86	329	EST02348	M85827	HFBC191	369	EST02381	M85870	HFBC257
199	EST02219	M85702	HFBC87	330	EST02349	M85828	HFBC192	370	EST02382	M85871	HFBC258
200	EST02220	M85703	HFBC88	331	EST02350	M85829	HFBC193	371	EST02383	M85872	HFBC259

SUBSTITUTE SHEET

SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone	SEQ ID	EST#	GB#	Clone
1314	EST02344	M85922	HFBC038	1446	EST02482	M85958	HFBC957	1446	EST02482	M85958	HFBC957
1315	EST02345	M85923	HFBC039	1447	EST02483	M85959	HFBC958	1447	EST02483	M85959	HFBC958
1316	EST02346	M85924	HFBC040	1448	EST02484	M85960	HFBC959	1448	EST02484	M85960	HFBC959
1317	EST02347	M85925	HFBC041	1449	EST02485	M85961	HFBC960	1449	EST02485	M85961	HFBC960
1318	EST02348	M85926	HFBC042	1450	EST02486	M85962	HFBC961	1450	EST02486	M85962	HFBC961
1319	EST02349	M85927	HFBC043	1451	EST02487	M85963	HFBC962	1451	EST02487	M85963	HFBC962
1320	EST02350	M85928	HFBC044	1452	EST02488	M85964	HFBC963	1452	EST02488	M85964	HFBC963
1321	EST02351	M85929	HFBC045	1453	EST02489	M85965	HFBC964	1453	EST02489	M85965	HFBC964
1322	EST02352	M85930	HFBC046	1454	EST02490	M85966	HFBC965	1454	EST02490	M85966	HFBC965
1323	EST02353	M85931	HFBC047	1455	EST02491	M85967	HFBC966	1455	EST02491	M85967	HFBC966
1324	EST02354	M85932	HFBC048	1456	EST02492	M85968	HFBC967	1456	EST02492	M85968	HFBC967
1325	EST02355	M85933	HFBC049	1457	EST02493	M85969	HFBC968	1457	EST02493	M85969	HFBC968
1326	EST02356	M85934	HFBC050	1458	EST02494	M85970	HFBC969	1458	EST02494	M85970	HFBC969
1327	EST02357	M85935	HFBC051	1459	EST02495	M85971	HFBC970	1459	EST02495	M85971	HFBC970
1328	EST02358	M85936	HFBC052	1460	EST02496	M85972	HFBC971	1460	EST02496	M85972	HFBC971
1329	EST02359	M85937	HFBC053	1461	EST02497	M85973	HFBC972	1461	EST02497	M85973	HFBC972
1330	EST02360	M85938	HFBC054	1462	EST02498	M85974	HFBC973	1462	EST02498	M85974	HFBC973
1331	EST02361	M85939	HFBC055	1463	EST02499	M85975	HFBC974	1463	EST02499	M85975	HFBC974
1332	EST02362	M85940	HFBC056	1464	EST02500	M85976	HFBC975	1464	EST02500	M85976	HFBC975
1333	EST02363	M85941	HFBC057	1465	EST02501	M85977	HFBC976	1465	EST02501	M85977	HFBC976
1334	EST02364	M85942	HFBC058	1466	EST02502	M85978	HFBC977	1466	EST02502	M85978	HFBC977
1335	EST02365	M85943	HFBC059	1467	EST02503	M85979	HFBC978	1467	EST02503	M85979	HFBC978
1336	EST02366	M85944	HFBC060	1468	EST02504	M85980	HFBC979	1468	EST02504	M85980	HFBC979
1337	EST02367	M85945	HFBC061	1469	EST02505	M85981	HFBC980	1469	EST02505	M85981	HFBC980
1338	EST02368	M85946	HFBC062	1470	EST02506	M85982	HFBC981	1470	EST02506	M85982	HFBC981
1339	EST02369	M85947	HFBC063	1471	EST02507	M85983	HFBC982	1471	EST02507	M85983	HFBC982
1340	EST02370	M85948	HFBC064	1472	EST02508	M85984	HFBC983	1472	EST02508	M85984	HFBC983
1341	EST02371	M85949	HFBC065	1473	EST02509	M85985	HFBC984	1473	EST02509	M85985	HFBC984
1342	EST02372	M85950	HFBC066	1474	EST02510	M85986	HFBC985	1474	EST02510	M85986	HFBC985
1343	EST02373	M85951	HFBC067	1475	EST02511	M85987	HFBC986	1475	EST02511	M85987	HFBC986
1344	EST02374	M85952	HFBC068	1476	EST02512	M85988	HFBC987	1476	EST02512	M85988	HFBC987
1345	EST02375	M85953	HFBC069	1477	EST02513	M85989	HFBC988	1477	EST02513	M85989	HFBC988
1346	EST02376	M85954	HFBC070	1478	EST02514	M85990	HFBC989	1478	EST02514	M85990	HFBC989
1347	EST02377	M85955	HFBC071	1479	EST02515	M85991	HFBC990	1479	EST02515	M85991	HFBC990
1348	EST02378	M85956	HFBC072	1480	EST02516	M85992	HFBC991	1480	EST02516	M85992	HFBC991
1349	EST02379	M85957	HFBC073	1481	EST02517	M85993	HFBC992	1481	EST02517	M85993	HFBC992
1350	EST02380	M85958	HFBC074	1482	EST02518	M85994	HFBC993	1482	EST02518	M85994	HFBC993
1351	EST02381	M85959	HFBC075	1483	EST02519	M85995	HFBC994	1483	EST02519	M85995	HFBC994
1352	EST02382	M85960	HFBC076	1484	EST02520	M85996	HFBC995	1484	EST02520	M85996	HFBC995
1353	EST02383	M85961	HFBC077	1485	EST02521	M85997	HFBC996	1485	EST02521	M85997	HFBC996
1354	EST02384	M85962	HFBC078	1486	EST02522	M85998	HFBC997	1486	EST02522	M85998	HFBC997
1355	EST02385	M85963	HFBC079	1487	EST02523	M85999	HFBC998	1487	EST02523	M85999	HFBC998
1356	EST02386	M85964	HFBC080	1488	EST02524	M86000	HFBC999	1488	EST02524	M86000	HFBC999
1357	EST02387	M85965	HFBC081	1489	EST02525	M86001	HFBC000	1489	EST02525	M86001	HFBC000
1358	EST02388	M85966	HFBC082	1490	EST02526	M86002	HFBC001	1490	EST02526	M86002	HFBC001
1359	EST02389	M85967	HFBC083	1491	EST02527	M86003	HFBC002	1491	EST02527	M86003	HFBC002
1360	EST02390	M85968	HFBC084	1492	EST02528	M86004	HFBC003	1492	EST02528	M86004	HFBC003
1361	EST02391	M85969	HFBC085	1493	EST02529	M86005	HFBC004	1493	EST02529	M86005	HFBC004
1362	EST02392	M85970	HFBC086	1494	EST02530	M86006	HFBC005	1494	EST02530	M86006	HFBC005
1363	EST02393	M85971	HFBC087	1495	EST02531	M86007	HFBC006	1495	EST02531	M86007	HFBC006
1364	EST02394	M85972	HFBC088	1496	EST02532	M86008	HFBC007	1496	EST02532	M86008	HFBC007
1365	EST02395	M85973	HFBC089	1497	EST02533	M86009	HFBC008	1497	EST02533	M86009	HFBC008
1366	EST02396	M85974	HFBC090	1498	EST02534	M86010	HFBC009	1498	EST02534	M86010	HFBC009
1367	EST02397	M85975	HFBC091	1499	EST02535	M86011	HFBC010	1499	EST02535	M86011	HFBC010
1368	EST02398	M85976	HFBC092	1500	EST02536	M86012	HFBC011	1500	EST02536	M86012	HFBC011
1369	EST02399	M85977	HFBC093	1501	EST02537	M86013	HFBC012	1501	EST02537	M86013	HFBC012
1370	EST02400	M85978	HFBC094	1502	EST02538	M86014	HFBC013	1502	EST02538	M86014	HFBC013
1371	EST02401	M85979	HFBC095								
1372	EST02402	M85980	HFBC096								
1373	EST02403	M85981	HFBC097								
1374	EST02404	M85982	HFBC098								
1375	EST02405	M85983	HFBC099								
1376	EST02406	M85984	HFBC100								
1377	EST02407	M85985	HFBC101								
1378	EST02408	M85986	HFBC102								
1379	EST02409	M85987	HFBC103								
1380	EST02410	M85988	HFBC104								

SUBSTITUTE SHEET

Seq ID	EST#	Gene	Seq ID	EST#	Gene	Seq ID	EST#	Gene	Seq ID	EST#	Gene
1503	EST02541	MB6016	1691	EST01577	HICM39	1504	EST02542	MB6017	1692	EST01578	HICM40
1505	EST02543	MB6018	1693	EST01579	HICM41	1506	EST02544	MB6019	1694	EST01580	HICM42
1507	EST02545	MB6020	1695	EST01581	HICM43	1508	EST02546	MB6021	1696	EST01582	HICM44
1509	EST02547	MB6022	1697	EST01583	HICM45	1510	EST02548	MB6023	1698	EST01584	HICM46
1511	EST02549	MB6024	1699	EST01585	HICM47	1512	EST02550	MB6025	1700	EST01586	HICM48
1513	EST02551	MB6026	1701	EST01587	HICM49	1514	EST02552	MB6027	1702	EST01588	HICM50
1515	EST02553	MB6028	1703	EST01589	HICM51	1516	EST02554	MB6029	1704	EST01590	HICM52
1517	EST02555	MB6030	1705	EST01591	HICM53	1518	EST02556	MB6031	1706	EST01592	HICM54
1519	EST02557	MB6032	1707	EST01593	HICM55	1520	EST02558	MB6033	1708	EST01594	HICM56
1521	EST02559	MB6034	1709	EST01595	HICM57	1522	EST02560	MB6035	1710	EST01596	HICM58
1523	EST02561	MB6036	1711	EST01597	HICM59	1524	EST02562	MB6037	1712	EST01598	HICM60
1525	EST02563	MB6038	1713	EST01599	HICM61	1526	EST02564	MB6039	1714	EST01600	HICM62
1527	EST02565	MB6039	1715	EST01601	HICM63	1528	EST02566	MB6040	1716	EST01602	HICM64
1529	EST02567	MB6041	1717	EST01603	HICM65	1530	EST02568	MB6042	1718	EST01604	HICM66
1531	EST02569	MB6043	1719	EST01605	HICM67	1532	EST02570	MB6044	1720	EST01606	HICM68
1533	EST02571	MB6045	1721	EST01607	HICM69	1534	EST02572	MB6046	1722	EST01608	HICM70
1535	EST02573	MB6047	1723	EST01609	HICM71	1536	EST02574	MB6048	1724	EST01610	HICM72
1537	EST02575	MB6049	1725	EST01611	HICM73	1538	EST02576	MB6050	1726	EST01612	HICM74
1539	EST02577	MB6051	1727	EST01613	HICM75	1540	EST02578	MB6052	1728	EST01614	HICM76
1541	EST02579	MB6053	1729	EST01615	HICM77	1542	EST02580	MB6054	1730	EST01616	HICM78
1543	EST02581	MB6055	1731	EST01617	HICM79	1544	EST02582	MB6056	1732	EST01618	HICM80
1545	EST02583	MB6057	1733	EST01619	HICM81	1546	EST02584	MB6058	1734	EST01620	HICM82
1547	EST02585	MB6059	1735	EST01621	HICM83	1548	EST02586	MB6060	1736	EST01622	HICM84
1549	EST02587	MB6061	1737	EST01623	HICM85	1550	EST02588	MB6062	1738	EST01624	HICM86
1551	EST02589	MB6063	1739	EST01625	HICM87	1552	EST02590	MB6064	1740	EST01626	HICM88
1553	EST02591	MB6065	1741	EST01627	HICM89	1554	EST02592	MB6066	1742	EST01628	HICM90
1555	EST02593	MB6067	1743	EST01629	HICM91	1556	EST02594	MB6068	1744	EST01630	HICM92
1557	EST02595	MB6069	1745	EST01631	HICM93	1558	EST02596	MB6070	1746	EST01632	HICM94
1559	EST02597	MB6071	1747	EST01633	HICM95	1560	EST02598	MB6072	1748	EST01634	HICM96
1561	EST02599	MB6073	1749	EST01635	HICM97	1562	EST02600	MB6074	1750	EST01636	HICM98
1563	EST02601	MB6075	1751	EST01637	HICM99	1564	EST02602	MB6076	1752	EST01638	HICM100
1565	EST02603	MB6077	1753	EST01639	HICM101	1566	EST02604	MB6078	1754	EST01640	HICM102
1567	EST02605	MB6079	1755	EST01641	HICM103	1568	EST02606	MB6080	1756	EST01642	HICM104
1569	EST02607	MB6081	1757	EST01643	HICM105	1570	EST02608	MB6082	1758	EST01644	HICM106
1571	EST02609	MB6083	1759	EST01645	HICM107	1572	EST02610	MB6084	1760	EST01646	HICM108
1573	EST02611	MB6085	1761	EST01647	HICM109	1574	EST02612	MB6086	1762	EST01648	HICM110
1575	EST02613	MB6087	1763	EST01649	HICM111	1576	EST02614	MB6088	1764	EST01650	HICM112
1577	EST02615	MB6089	1765	EST01651	HICM113	1578	EST02616	MB6090	1766	EST01652	HICM114
1579	EST02617	MB6091	1767	EST01653	HICM115	1580	EST02618	MB6092	1768	EST01654	HICM116
1581	EST02619	MB6093	1769	EST01655	HICM117	1582	EST02620	MB6094	1770	EST01656	HICM118
1583	EST02621	MB6095	1771	EST01657	HICM119	1584	EST02622	MB6096	1772	EST01658	HICM120
1585	EST02623	MB6097	1773	EST01659	HICM121	1586	EST02624	MB6098	1774	EST01660	HICM122
1587	EST02625	MB6099	1775	EST01661	HICM123	1588	EST02626	MB6100	1776	EST01662	HICM124
1589	EST02627	MB6101	1777	EST01663	HICM125	1590	EST02628	MB6102	1778	EST01664	HICM126
1591	EST02629	MB6103	1779	EST01665	HICM127	1592	EST02630	MB6104	1780	EST01666	HICM128
1593	EST02631	MB6105	1781	EST01667	HICM129	1594	EST02632	MB6106	1782	EST01668	HICM130
1595	EST02633	MB6107	1783	EST01669	HICM131	1596	EST02634	MB6108	1784	EST01670	HICM132
1597	EST02635	MB6109	1785	EST01671	HICM133	1598	EST02636	MB6110	1786	EST01672	HICM134
1599	EST02637	MB6111	1787	EST01673	HICM135	1600	EST02638	MB6112	1788	EST01674	HICM136
1601	EST02639	MB6113	1789	EST01675	HICM137	1602	EST02640	MB6114	1790	EST01676	HICM138
1603	EST02641	MB6115	1791	EST01677	HICM139	1604	EST02642	MB6116	1792	EST01678	HICM140
1605	EST02643	MB6117	1793	EST01679	HICM141	1606	EST02644	MB6118	1794	EST01680	HICM142
1607	EST02645	MB6119	1795	EST01681	HICM143	1608	EST02646	MB6120	1796	EST01682	HICM144
1609	EST02647	MB6121	1797	EST01683	HICM145	1610	EST02648	MB6122	1798	EST01684	HICM146
1611	EST02649	MB6123	1799	EST01685	HICM147	1612	EST02650	MB6124	1800	EST01686	HICM148
1613	EST02651	MB6125	1801	EST01687	HICM149	1614	EST02652	MB6126	1802	EST01688	HICM150
1615	EST02653	MB6127	1803	EST01689	HICM151	1616	EST02654	MB6128	1804	EST01690	HICM152
1617	EST02655	MB6129	1805	EST01691	HICM153	1618	EST02656	MB6130	1806	EST01692	HICM154
1619	EST02657	MB6131	1807	EST01693	HICM155	1620	EST02658	MB6132	1808	EST01694	HICM156
1621	EST02659	MB6133	1809	EST01695	HICM157	1622	EST02660	MB6134	1810	EST01696	HICM158
1623	EST02661	MB6135	1811	EST01697	HICM159	1624	EST02662	MB6136	1812	EST01698	HICM160
1625	EST02663	MB6137	1813	EST01699	HICM161	1626	EST02664	MB6138	1814	EST01700	HICM162
1627	EST02665	MB6139	1815	EST01701	HICM163	1628	EST02666	MB6140	1816	EST01702	HICM164
1629	EST02667	MB6141	1817	EST01703	HICM165	1630	EST02668	MB6142	1818	EST01704	HICM166
1631	EST02669	MB6143	1819	EST01705	HICM167	1632	EST02670	MB6144	1820	EST01706	HICM168
1633	EST02671	MB6145	1821	EST01707	HICM169	1634	EST02672	MB6146	1822	EST01708	HICM170
1635	EST02673	MB6147	1823	EST01709	HICM171	1636	EST02674	MB6148	1824	EST01710	HICM172
1637	EST02675	MB6149	1825	EST01711	HICM173	1638	EST02676	MB6150	1826	EST01712	HICM174
1639	EST02677	MB6151	1827	EST01713	HICM175	1640	EST02678	MB6152	1828	EST01714	HICM176
1641	EST02679	MB6153	1829	EST01715	HICM177	1642	EST02680	MB6154	1830	EST01716	HICM178
1643	EST02681	MB6155	1831	EST01717	HICM179	1644	EST02682	MB6156	1832	EST01718	HICM180
1645	EST02683	MB6157	1833	EST01719	HICM181	1646	EST02684	MB6158	1834	EST01720	HICM182
1647	EST02685	MB6159	1835	EST01721	HICM183	1648	EST02686	MB6160	1836	EST01722	HICM184
1649	EST02687	MB6161	1837	EST01723	HICM185	1650	EST02688	MB6162	1838	EST01724	HICM186
1651	EST02689	MB6163	1839	EST01725	HICM187	1652	EST02690	MB6164	1840	EST01726	HICM188
1653	EST02691	MB6165	1841	EST01727	HICM189	1654	EST02692	MB6166	1842	EST01728	HICM190
1655	EST02693	MB6167	1843	EST01729	HICM191	1656	EST02694	MB6168	1844	EST01730	HICM192
1657	EST02695	MB6169	1845	EST01731	HICM193	1658	EST02696	MB6170	1846	EST01732	HICM194
1659	EST02697	MB6171	1847	EST01733	HICM195	1660	EST02698	MB6172	1848	EST01734	HICM196
1661	EST02699	MB6173	1849	EST01735	HICM197	1662	EST02700	MB6174	1850	EST01736	HICM198
1663	EST02701	MB6175	1851	EST01737	HICM199	1664	EST02702	MB6176	1852	EST01738	HICM200
1665	EST02703	MB6177	1853	EST01739	HICM201	1666	EST02704	MB6178	1854	EST01740	HICM202
1667	EST02705	MB6179	1855	EST01741	HICM203	1668	EST02706	MB6180	1856	EST01742	HICM204
1669	EST02707	MB6181	1857	EST01743	HICM205	1670	EST02708	MB6182	1858	EST01744	HICM206
1671	EST02709	MB6183	1859	EST01745	HICM207	1672	EST02710	MB6184	1860	EST01746	HICM208
1673	EST02711	MB6185	1861	EST01747	HICM209	1674	EST02712	MB6186	1862	EST01748	HICM210
1675	EST02713	MB6187	1863	EST01749	HICM211	1676	EST02714	MB6188	1864	EST01750	HICM212
1677	EST02715	MB6189	1865	EST01751	HICM213	1678	EST02716	MB6190	1866	EST01752	HICM214
1679	EST02717	MB6191	1867	EST01753	HICM215	1680	EST02718	MB6192	1868	EST01754	HICM216
1681	EST02719	MB6193	1869	EST01755	HICM217	1682	EST02720	MB6194	1870	EST01756	HICM218
1683	EST02721	MB6195	1871	EST01757	HICM219	1684	EST02722	MB6196	1872	EST01758	HICM220
1685	EST02723	MB6197	1873	EST01759	HICM221	1686	EST02724	MB6198	1874	EST01760	HICM222
1687	EST02725	MB6199	1875	EST01761	HICM223	1688	EST02726	MB6200	1876	EST01762	HICM224
1689	EST02727	MB6201	1877	EST01763	HICM225	1690	EST02728	MB6202	1878	EST01764	HICM226
1691	EST02729	MB6203	1879	EST01765	HICM227	1692	EST02730	MB6204	1880	EST01766	HICM228
1693	EST02731	MB6205	1881	EST01767	HICM229	1694	EST02732	MB6206	1882	EST01768	HICM230
1695	EST02733	MB6207	1883	EST01769	HICM231	1696	EST02734	MB6208	1884	EST01770	HICM

SEQ. ID	EST#	GS#	Clone	SEQ. ID	EST#	GS#	Clone	SEQ. ID	EST#	GS#	Clone
1692	EST00847	W78699	HHCHC42	1824	EST01617	W78030	HHCHG04	1824	EST01617	W78030	HHCHG04
1693	EST00848	W78700	HHCHC43	1825	EST00955	W78807	HHCHG07	1825	EST00955	W78807	HHCHG07
1694	EST00849	W78701	HHCHC44	1826	EST00956	W78808	HHCHG08	1826	EST00956	W78808	HHCHG08
1695	EST00850	W78702	HHCHC45	1827	EST01618	W78031	HHCHG10	1827	EST01618	W78031	HHCHG10
1696	EST00851	W78703	HHCHC46	1828	EST00957	W78809	HHCHG11	1828	EST00957	W78809	HHCHG11
1697	EST00852	W78704	HHCHC47	1829	EST00958	W78810	HHCHG12	1829	EST00958	W78810	HHCHG12
1698	EST02685	W86154	HHCHC48	1830	EST00959	W78811	HHCHG13	1830	EST00959	W78811	HHCHG13
1699	EST02686	W86155	HHCHC49	1831	EST01619	W78032	HHCHG15	1831	EST01619	W78032	HHCHG15
1700	EST00853	W78705	HHCHC50	1832	EST00960	W78812	HHCHG16	1832	EST00960	W78812	HHCHG16
1701	EST00854	W78706	HHCHC51	1833	EST00961	W78813	HHCHG17	1833	EST00961	W78813	HHCHG17
1702	EST00855	W78707	HHCHC52	1834	EST00962	W78814	HHCHG18	1834	EST00962	W78814	HHCHG18
1703	EST01580	W78708	HHCHC53	1835	EST01620	W78033	HHCHG21	1835	EST01620	W78033	HHCHG21
1704	EST00856	W78709	HHCHC54	1836	EST00963	W78815	HHCHG22	1836	EST00963	W78815	HHCHG22
1705	EST00857	W78710	HHCHC55	1837	EST00964	W78816	HHCHG23	1837	EST00964	W78816	HHCHG23
1706	EST01581	W78711	HHCHC56	1838	EST00965	W78817	HHCHG24	1838	EST00965	W78817	HHCHG24
1707	EST00858	W78712	HHCHC57	1839	EST00966	W78818	HHCHG25	1839	EST00966	W78818	HHCHG25
1708	EST00859	W78713	HHCHC58	1840	EST00967	W78819	HHCHG26	1840	EST00967	W78819	HHCHG26
1709	EST00860	W78714	HHCHC59	1841	EST00968	W78820	HHCHG27	1841	EST00968	W78820	HHCHG27
1710	EST00861	W78715	HHCHC60	1842	EST00969	W78821	HHCHG28	1842	EST00969	W78821	HHCHG28
1711	EST00862	W78716	HHCHC61	1843	EST00970	W78822	HHCHG29	1843	EST00970	W78822	HHCHG29
1712	EST00863	W78717	HHCHC62	1844	EST00971	W78823	HHCHG30	1844	EST00971	W78823	HHCHG30
1713	EST00864	W78718	HHCHC63	1845	EST00972	W78824	HHCHG31	1845	EST00972	W78824	HHCHG31
1714	EST01582	W78719	HHCHC64	1846	EST00973	W78825	HHCHG32	1846	EST00973	W78825	HHCHG32
1715	EST01583	W78720	HHCHC65	1847	EST00974	W78826	HHCHG33	1847	EST00974	W78826	HHCHG33
1716	EST00865	W78721	HHCHC66	1848	EST00975	W78827	HHCHG34	1848	EST00975	W78827	HHCHG34
1717	EST00866	W78722	HHCHC67	1849	EST01624	W78037	HHCHG35	1849	EST01624	W78037	HHCHG35
1718	EST00867	W78723	HHCHC68	1850	EST00976	W78828	HHCHG36	1850	EST00976	W78828	HHCHG36
1719	EST00868	W78724	HHCHC69	1851	EST01625	W78038	HHCHG37	1851	EST01625	W78038	HHCHG37
1720	EST00869	W78725	HHCHC70	1852	EST00977	W78829	HHCHG38	1852	EST00977	W78829	HHCHG38
1721	EST00870	W78726	HHCHC71	1853	EST00978	W78830	HHCHG39	1853	EST00978	W78830	HHCHG39
1722	EST00871	W78727	HHCHC72	1854	EST01627	W78040	HHCHG40	1854	EST01627	W78040	HHCHG40
1723	EST00872	W78728	HHCHC73	1855	EST00979	W78831	HHCHG41	1855	EST00979	W78831	HHCHG41
1724	EST00873	W78729	HHCHC74	1856	EST00980	W78832	HHCHG42	1856	EST00980	W78832	HHCHG42
1725	EST01584	W78730	HHCHC75	1857	EST00981	W78833	HHCHG43	1857	EST00981	W78833	HHCHG43
1726	EST00874	W78731	HHCHC76	1858	EST00982	W78834	HHCHG44	1858	EST00982	W78834	HHCHG44
1727	EST00875	W78732	HHCHC77	1859	EST00983	W78835	HHCHG45	1859	EST00983	W78835	HHCHG45
1728	EST00876	W78733	HHCHC78	1860	EST00984	W78836	HHCHG46	1860	EST00984	W78836	HHCHG46
1729	EST00877	W78734	HHCHC79	1861	EST00985	W78837	HHCHG47	1861	EST00985	W78837	HHCHG47
1730	EST00878	W78735	HHCHC80	1862	EST00986	W78838	HHCHG48	1862	EST00986	W78838	HHCHG48
1731	EST01590	W78736	HHCHC81	1863	EST00987	W78839	HHCHG49	1863	EST00987	W78839	HHCHG49
1732	EST01591	W78737	HHCHC82	1864	EST00988	W78840	HHCHG50	1864	EST00988	W78840	HHCHG50
1733	EST00880	W78738	HHCHC83	1865	EST00989	W78841	HHCHG51	1865	EST00989	W78841	HHCHG51
1734	EST00881	W78739	HHCHC84	1866	EST00990	W78842	HHCHG52	1866	EST00990	W78842	HHCHG52
1735	EST01592	W78740	HHCHC85	1867	EST00991	W78843	HHCHG53	1867	EST00991	W78843	HHCHG53
1736	EST00882	W78741	HHCHC86	1868	EST00992	W78844	HHCHG54	1868	EST00992	W78844	HHCHG54
1737	EST00883	W78742	HHCHC87	1869	EST00993	W78845	HHCHG55	1869	EST00993	W78845	HHCHG55
1738	EST00884	W78743	HHCHC88	1870	EST00994	W78846	HHCHG56	1870	EST00994	W78846	HHCHG56
1739	EST02687	W86156	HHCHC89	1871	EST00995	W78847	HHCHG57	1871	EST00995	W78847	HHCHG57
1740	EST00885	W78744	HHCHC90	1872	EST01630	W78043	HHCHG58	1872	EST01630	W78043	HHCHG58
1741	EST00886	W78745	HHCHC91	1873	EST00996	W78848	HHCHG59	1873	EST00996	W78848	HHCHG59
1742	EST00887	W78746	HHCHC92	1874	EST01631	W78044	HHCHG60	1874	EST01631	W78044	HHCHG60
1743	EST00888	W78747	HHCHC93	1875	EST00997	W78849	HHCHG61	1875	EST00997	W78849	HHCHG61
1744	EST00889	W78748	HHCHC94	1876	EST00998	W78850	HHCHG62	1876	EST00998	W78850	HHCHG62
1745	EST00890	W78749	HHCHC95	1877	EST00999	W78851	HHCHG63	1877	EST00999	W78851	HHCHG63
1746	EST00891	W78750	HHCHC96	1878	EST01633	W78046	HHCHG64	1878	EST01633	W78046	HHCHG64
1747	EST00892	W78751	HHCHC97	1879	EST01634	W78047	HHCHG65	1879	EST01634	W78047	HHCHG65
1748	EST00893	W78752	HHCHC98								
1749	EST01593	W78753	HHCHC99								
1750	EST00894	W78754	HHCHC00								
1751	EST00895	W78755	HHCHC01								
1752	EST00896	W78756	HHCHC02								
1753	EST00897	W78757	HHCHC03								
1754	EST00898	W78758	HHCHC04								
1755	EST00899	W78759	HHCHC05								
1756	EST00900	W78760	HHCHC06								
1757	EST01594	W78761	HHCHC07								

SUBSTITUTE SHEET

SEQ ID	EST#	Clone	CS#	Clone	EST#	SEQ ID	EST#	Clone	CS#	Clone	EST#	SEQ ID
1881	EST01000	HHCPH48	W78852	HHCPH48	EST01053	1947	EST01053	HHCPH48	W78904	HHCPH48	EST01107	2013
1882	EST01001	HHCPH49	W78853	HHCPH49	EST01054	1948	EST01054	HHCPH49	W78905	HHCPH49	EST01108	2014
1883	EST01002	HHCPH50	W78854	HHCPH50	EST01055	1949	EST01055	HHCPH50	W78906	HHCPH50	EST01109	2015
1884	EST01003	HHCPH51	W78855	HHCPH51	EST01056	1950	EST01056	HHCPH51	W78907	HHCPH51	EST01110	2016
1885	EST01004	HHCPH52	W78856	HHCPH52	EST01057	1951	EST01057	HHCPH52	W78908	HHCPH52	EST01111	2017
1886	EST01005	HHCPH53	W78857	HHCPH53	EST01058	1952	EST01058	HHCPH53	W78909	HHCPH53	EST01112	2018
1887	EST01006	HHCPH54	W78858	HHCPH54	EST01059	1953	EST01059	HHCPH54	W78910	HHCPH54	EST01113	2019
1888	EST01007	HHCPH55	W78859	HHCPH55	EST01060	1954	EST01060	HHCPH55	W78911	HHCPH55	EST01114	2020
1889	EST01008	HHCPH56	W78860	HHCPH56	EST01061	1955	EST01061	HHCPH56	W78912	HHCPH56	EST01115	2021
1890	EST01009	HHCPH57	W78861	HHCPH57	EST01062	1956	EST01062	HHCPH57	W78913	HHCPH57	EST01116	2022
1891	EST01010	HHCPH58	W78862	HHCPH58	EST01063	1957	EST01063	HHCPH58	W78914	HHCPH58	EST01117	2023
1892	EST01011	HHCPH59	W78863	HHCPH59	EST01064	1958	EST01064	HHCPH59	W78915	HHCPH59	EST01118	2024
1893	EST01012	HHCPH60	W78864	HHCPH60	EST01065	1959	EST01065	HHCPH60	W78916	HHCPH60	EST01119	2025
1894	EST01013	HHCPH61	W78865	HHCPH61	EST01066	1960	EST01066	HHCPH61	W78917	HHCPH61	EST01120	2026
1895	EST01014	HHCPH62	W78866	HHCPH62	EST01067	1961	EST01067	HHCPH62	W78918	HHCPH62	EST01121	2027
1896	EST01015	HHCPH63	W78867	HHCPH63	EST01068	1962	EST01068	HHCPH63	W78919	HHCPH63	EST01122	2028
1897	EST01016	HHCPH64	W78868	HHCPH64	EST01069	1963	EST01069	HHCPH64	W78920	HHCPH64	EST01123	2029
1898	EST01017	HHCPH65	W78869	HHCPH65	EST01070	1964	EST01070	HHCPH65	W78921	HHCPH65	EST01124	2030
1899	EST01018	HHCPH66	W78870	HHCPH66	EST01071	1965	EST01071	HHCPH66	W78922	HHCPH66	EST01125	2031
1900	EST01019	HHCPH67	W78871	HHCPH67	EST01072	1966	EST01072	HHCPH67	W78923	HHCPH67	EST01126	2032
1901	EST01020	HHCPH68	W78872	HHCPH68	EST01073	1967	EST01073	HHCPH68	W78924	HHCPH68	EST01127	2033
1902	EST01021	HHCPH69	W78873	HHCPH69	EST01074	1968	EST01074	HHCPH69	W78925	HHCPH69	EST01128	2034
1903	EST01022	HHCPH70	W78874	HHCPH70	EST01075	1969	EST01075	HHCPH70	W78926	HHCPH70	EST01129	2035
1904	EST01023	HHCPH71	W78875	HHCPH71	EST01076	1970	EST01076	HHCPH71	W78927	HHCPH71	EST01130	2036
1905	EST01024	HHCPH72	W78876	HHCPH72	EST01077	1971	EST01077	HHCPH72	W78928	HHCPH72	EST01131	2037
1906	EST01025	HHCPH73	W78877	HHCPH73	EST01078	1972	EST01078	HHCPH73	W78929	HHCPH73	EST01132	2038
1907	EST01026	HHCPH74	W78878	HHCPH74	EST01079	1973	EST01079	HHCPH74	W78930	HHCPH74	EST01133	2039
1908	EST01027	HHCPH75	W78879	HHCPH75	EST01080	1974	EST01080	HHCPH75	W78931	HHCPH75	EST01134	2040
1909	EST01028	HHCPH76	W78880	HHCPH76	EST01081	1975	EST01081	HHCPH76	W78932	HHCPH76	EST01135	2041
1910	EST01029	HHCPH77	W78881	HHCPH77	EST01082	1976	EST01082	HHCPH77	W78933	HHCPH77	EST01136	2042
1911	EST01030	HHCPH78	W78882	HHCPH78	EST01083	1977	EST01083	HHCPH78	W78934	HHCPH78	EST01137	2043
1912	EST01031	HHCPH79	W78883	HHCPH79	EST01084	1978	EST01084	HHCPH79	W78935	HHCPH79	EST01138	2044
1913	EST01032	HHCPH80	W78884	HHCPH80	EST01085	1979	EST01085	HHCPH80	W78936	HHCPH80	EST01139	2045
1914	EST01033	HHCPH81	W78885	HHCPH81	EST01086	1980	EST01086	HHCPH81	W78937	HHCPH81	EST01140	2046
1915	EST01034	HHCPH82	W78886	HHCPH82	EST01087	1981	EST01087	HHCPH82	W78938	HHCPH82	EST01141	2047
1916	EST01035	HHCPH83	W78887	HHCPH83	EST01088	1982	EST01088	HHCPH83	W78939	HHCPH83	EST01142	2048
1917	EST01036	HHCPH84	W78888	HHCPH84	EST01089	1983	EST01089	HHCPH84	W78940	HHCPH84	EST01143	2049
1918	EST01037	HHCPH85	W78889	HHCPH85	EST01090	1984	EST01090	HHCPH85	W78941	HHCPH85	EST01144	2050
1919	EST01038	HHCPH86	W78890	HHCPH86	EST01091	1985	EST01091	HHCPH86	W78942	HHCPH86	EST01145	2051
1920	EST01039	HHCPH87	W78891	HHCPH87	EST01092	1986	EST01092	HHCPH87	W78943	HHCPH87	EST01146	2052
1921	EST01040	HHCPH88	W78892	HHCPH88	EST01093	1987	EST01093	HHCPH88	W78944	HHCPH88	EST01147	2053
1922	EST01041	HHCPH89	W78893	HHCPH89	EST01094	1988	EST01094	HHCPH89	W78945	HHCPH89	EST01148	2054
1923	EST01042	HHCPH90	W78894	HHCPH90	EST01095	1989	EST01095	HHCPH90	W78946	HHCPH90	EST01149	2055
1924	EST01043	HHCPH91	W78895	HHCPH91	EST01096	1990	EST01096	HHCPH91	W78947	HHCPH91	EST01150	2056
1925	EST01044	HHCPH92	W78896	HHCPH92	EST01097	1991	EST01097	HHCPH92	W78948	HHCPH92	EST01151	2057
1926	EST01045	HHCPH93	W78897	HHCPH93	EST01098	1992	EST01098	HHCPH93	W78949	HHCPH93	EST01152	2058
1927	EST01046	HHCPH94	W78898	HHCPH94	EST01099	1993	EST01099	HHCPH94	W78950	HHCPH94	EST01153	2059
1928	EST01047	HHCPH95	W78899	HHCPH95	EST01100	1994	EST01100	HHCPH95	W78951	HHCPH95	EST01154	2060
1929	EST01048	HHCPH96	W78900	HHCPH96	EST01101	1995	EST01101	HHCPH96	W78952	HHCPH96	EST01155	2061
1930	EST01049	HHCPH97	W78901	HHCPH97	EST01102	1996	EST01102	HHCPH97	W78953	HHCPH97	EST01156	2062
1931	EST01050	HHCPH98	W78902	HHCPH98	EST01103	1997	EST01103	HHCPH98	W78954	HHCPH98	EST01157	2063
1932	EST01051	HHCPH99	W78903	HHCPH99	EST01104	1998	EST01104	HHCPH99	W78955	HHCPH99	EST01158	2064
1933	EST01052	HHCPH100	W78904	HHCPH100	EST01105	1999	EST01105	HHCPH100	W78956	HHCPH100	EST01159	2065
1934	EST01053	HHCPH101	W78905	HHCPH101	EST01106	2000	EST01106	HHCPH101	W78957	HHCPH101	EST01160	2066
1935	EST01054	HHCPH102	W78906	HHCPH102	EST01107	2001	EST01107	HHCPH102	W78958	HHCPH102	EST01161	2067
1936	EST01055	HHCPH103	W78907	HHCPH103	EST01108	2002	EST01108	HHCPH103	W78959	HHCPH103	EST01162	2068
1937	EST01056	HHCPH104	W78908	HHCPH104	EST01109	2003	EST01109	HHCPH104	W78960	HHCPH104	EST01163	2069
1938	EST01057	HHCPH105	W78909	HHCPH105	EST01110	2004	EST01110	HHCPH105	W78961	HHCPH105	EST01164	2070
1939	EST01058	HHCPH106	W78910	HHCPH106	EST01111	2005	EST01111	HHCPH106	W78962	HHCPH106	EST01165	2071
1940	EST01059	HHCPH107	W78911	HHCPH107	EST01112	2006	EST01112	HHCPH107	W78963	HHCPH107	EST01166	2072
1941	EST01060	HHCPH108	W78912	HHCPH108	EST01113	2007	EST01113	HHCPH108	W78964	HHCPH108	EST01167	2073
1942	EST01061	HHCPH109	W78913	HHCPH109	EST01114	2008	EST01114	HHCPH109	W78965	HHCPH109	EST01168	2074
1943	EST01062	HHCPH110	W78914	HHCPH110	EST01115	2009	EST01115	HHCPH110	W78966	HHCPH110	EST01169	2075
1944	EST01063	HHCPH111	W78915	HHCPH111	EST01116	2010	EST01116	HHCPH111	W78967	HHCPH111	EST01170	2076
1945	EST01064	HHCPH112	W78916	HHCPH112	EST01117	2011	EST01117	HHCPH112	W78968	HHCPH112	EST01171	2077
1946	EST01065	HHCPH113	W78917	HHCPH113	EST01118	2012	EST01118	HHCPH113	W78969	HHCPH113	EST01172	2078

Seq. ID	EST#	GB#	Clone	Seq. ID	EST#	GB#	Clone	Seq. ID	EST#	GB#	Clone	Seq. ID	EST#	GB#	Clone
2070	EST01151	M79003	HHCPH24	2136	EST01210	M79062	HHCPJ33	2202	EST01732	M79121	HHCPN163	2257	EST01754	M79155	HHCPN50
2071	EST01152	M78106	HHCPH32	2137	EST01211	M79063	HHCPJ40	2203	EST01264	M79116	HHCPN165	2258	EST01268	M79120	HHCPN50
2072	EST01153	M79004	HHCPH33	2138	EST011715	M78124	HHCPJ42	2204	EST01265	M79117	HHCPN166	2259	EST01269	M79121	HHCPN50
2073	EST01154	M78107	HHCPH34	2139	EST011716	M79064	HHCPJ44	2205	EST01266	M79118	HHCPN167	2260	EST01270	M79122	HHCPN50
2074	EST01155	M79005	HHCPH35	2140	EST01212	M79065	HHCPJ50	2206	EST01267	M79119	HHCPN168	2261	EST01271	M79123	HHCPN50
2075	EST01156	M78108	HHCPH36	2141	EST01213	M79066	HHCPJ54	2207	EST01268	M79120	HHCPN169	2262	EST01272	M79124	HHCPN50
2076	EST01157	M79006	HHCPH37	2142	EST01214	M79067	HHCPJ56	2208	EST01269	M79121	HHCPN170	2263	EST01273	M79125	HHCPN50
2077	EST01158	M79007	HHCPH38	2143	EST01215	M79068	HHCPJ58	2209	EST01270	M79122	HHCPN171	2264	EST01274	M79126	HHCPN50
2078	EST01159	M79008	HHCPH39	2144	EST01216	M79069	HHCPJ59	2210	EST01271	M79123	HHCPN172	2265	EST01275	M79127	HHCPN50
2079	EST01160	M79009	HHCPH40	2145	EST01217	M79070	HHCPJ60	2211	EST01272	M79124	HHCPN173	2266	EST01276	M79128	HHCPN50
2080	EST01161	M79010	HHCPH41	2146	EST01218	M79071	HHCPJ62	2212	EST01273	M79125	HHCPN174	2267	EST01277	M79129	HHCPN50
2081	EST01162	M79011	HHCPH42	2147	EST01219	M79072	HHCPJ64	2213	EST01274	M79126	HHCPN175	2268	EST01278	M79130	HHCPN50
2082	EST01163	M79012	HHCPH43	2148	EST01220	M79073	HHCPJ66	2214	EST01275	M79127	HHCPN176	2269	EST01279	M79131	HHCPN50
2083	EST01164	M79013	HHCPH44	2149	EST01221	M79074	HHCPJ68	2215	EST01276	M79128	HHCPN177	2270	EST01280	M79132	HHCPN50
2084	EST01165	M79014	HHCPH45	2150	EST01222	M79075	HHCPJ70	2216	EST01277	M79129	HHCPN178	2271	EST01281	M79133	HHCPN50
2085	EST01166	M79015	HHCPH46	2151	EST01223	M79076	HHCPJ72	2217	EST01278	M79130	HHCPN179	2272	EST01282	M79134	HHCPN50
2086	EST01167	M79016	HHCPH47	2152	EST01224	M79077	HHCPJ74	2218	EST01279	M79131	HHCPN180	2273	EST01283	M79135	HHCPN50
2087	EST01168	M79017	HHCPH48	2153	EST01225	M79078	HHCPJ76	2219	EST01280	M79132	HHCPN181	2274	EST01284	M79136	HHCPN50
2088	EST01169	M79018	HHCPH49	2154	EST01226	M79079	HHCPJ78	2220	EST01281	M79133	HHCPN182	2275	EST01285	M79137	HHCPN50
2089	EST01170	M79019	HHCPH50	2155	EST01227	M79080	HHCPJ80	2221	EST01282	M79134	HHCPN183	2276	EST01286	M79138	HHCPN50
2090	EST01171	M79020	HHCPH51	2156	EST01228	M79081	HHCPJ82	2222	EST01283	M79135	HHCPN184	2277	EST01287	M79139	HHCPN50
2091	EST01172	M79021	HHCPH52	2157	EST01229	M79082	HHCPJ84	2223	EST01284	M79136	HHCPN185	2278	EST01288	M79140	HHCPN50
2092	EST01173	M79022	HHCPH53	2158	EST01230	M79083	HHCPJ86	2224	EST01285	M79137	HHCPN186	2279	EST01289	M79141	HHCPN50
2093	EST01174	M79023	HHCPH54	2159	EST01231	M79084	HHCPJ88	2225	EST01286	M79138	HHCPN187	2280	EST01290	M79142	HHCPN50
2094	EST01175	M79024	HHCPH55	2160	EST01232	M79085	HHCPJ90	2226	EST01287	M79139	HHCPN188	2281	EST01291	M79143	HHCPN50
2095	EST01176	M79025	HHCPH56	2161	EST01233	M79086	HHCPJ92	2227	EST01288	M79140	HHCPN189	2282	EST01292	M79144	HHCPN50
2096	EST01177	M79026	HHCPH57	2162	EST01234	M79087	HHCPJ94	2228	EST01289	M79141	HHCPN190	2283	EST01293	M79145	HHCPN50
2097	EST01178	M79027	HHCPH58	2163	EST01235	M79088	HHCPJ96	2229	EST01290	M79142	HHCPN191	2284	EST01294	M79146	HHCPN50
2098	EST01179	M79028	HHCPH59	2164	EST01236	M79089	HHCPJ98	2230	EST01291	M79143	HHCPN192	2285	EST01295	M79147	HHCPN50
2099	EST01180	M79029	HHCPH60	2165	EST01237	M79090	HHCPJ100	2231	EST01292	M79144	HHCPN193	2286	EST01296	M79148	HHCPN50
2100	EST01181	M79030	HHCPH61	2166	EST01238	M79091	HHCPJ102	2232	EST01293	M79145	HHCPN194	2287	EST01297	M79149	HHCPN50
2101	EST01182	M79031	HHCPH62	2167	EST01239	M79092	HHCPJ104	2233	EST01294	M79146	HHCPN195	2288	EST01298	M79150	HHCPN50
2102	EST01183	M79032	HHCPH63	2168	EST01240	M79093	HHCPJ106	2234	EST01295	M79147	HHCPN196	2289	EST01299	M79151	HHCPN50
2103	EST01184	M79033	HHCPH64	2169	EST01241	M79094	HHCPJ108	2235	EST01296	M79148	HHCPN197	2290	EST01300	M79152	HHCPN50
2104	EST01185	M79034	HHCPH65	2170	EST01242	M79095	HHCPJ110	2236	EST01297	M79149	HHCPN198	2291	EST01301	M79153	HHCPN50
2105	EST01186	M79035	HHCPH66	2171	EST01243	M79096	HHCPJ112	2237	EST01298	M79150	HHCPN199	2292	EST01302	M79154	HHCPN50
2106	EST01187	M79036	HHCPH67	2172	EST01244	M79097	HHCPJ114	2238	EST01299	M79151	HHCPN200	2293	EST01303	M79155	HHCPN50
2107	EST01188	M79037	HHCPH68	2173	EST01245	M79098	HHCPJ116	2239	EST01300	M79152	HHCPN201	2294	EST01304	M79156	HHCPN50
2108	EST01189	M79038	HHCPH69	2174	EST01246	M79099	HHCPJ118	2240	EST01301	M79153	HHCPN202	2295	EST01305	M79157	HHCPN50
2109	EST01190	M79039	HHCPH70	2175	EST01247	M79100	HHCPJ120	2241	EST01302	M79154	HHCPN203	2296	EST01306	M79158	HHCPN50
2110	EST01191	M79040	HHCPH71	2176	EST01248	M79101	HHCPJ122	2242	EST01303	M79155	HHCPN204	2297	EST01307	M79159	HHCPN50
2111	EST01192	M79041	HHCPH72	2177	EST01249	M79102	HHCPJ124	2243	EST01304	M79160	HHCPN205	2298	EST01308	M79161	HHCPN50
2112	EST01193	M79042	HHCPH73	2178	EST01250	M79103	HHCPJ126	2244	EST01305	M79161	HHCPN206	2299	EST01309	M79162	HHCPN50
2113	EST01194	M79043	HHCPH74	2179	EST01251	M79104	HHCPJ128	2245	EST01306	M79162	HHCPN207	2300	EST01310	M79163	HHCPN50
2114	EST01195	M79044	HHCPH75	2180	EST01252	M79105	HHCPJ130	2246	EST01307	M79163	HHCPN208	2301	EST01311	M79164	HHCPN50
2115	EST01196	M79045	HHCPH76	2181	EST01253	M79106	HHCPJ132	2247	EST01308	M79164	HHCPN209	2302	EST01312	M79165	HHCPN50
2116	EST01197	M79046	HHCPH77	2182	EST01254	M79107	HHCPJ134	2248	EST01309	M79165	HHCPN210	2303	EST01313	M79166	HHCPN50
2117	EST01198	M79047	HHCPH78	2183	EST01255	M79108	HHCPJ136	2249	EST01310	M79166	HHCPN211	2304	EST01314	M79167	HHCPN50
2118	EST01199	M79048	HHCPH79	2184	EST01256	M79109	HHCPJ138	2250	EST01311	M79167	HHCPN212	2305	EST01315	M79168	HHCPN50
2119	EST01200	M79049	HHCPH80	2185	EST01257	M79110	HHCPJ140	2251	EST01312	M79168	HHCPN213	2306	EST01316	M79169	HHCPN50
2120	EST01201	M79050	HHCPH81	2186	EST01258	M79111	HHCPJ142	2252	EST01313	M79169	HHCPN214	2307	EST01317	M79170	HHCPN50
2121	EST01202	M79051	HHCPH82	2187	EST01259	M79112	HHCPJ144	2253	EST01314	M79170	HHCPN215	2308	EST01318	M79171	HHCPN50
2122	EST01203	M79052	HHCPH83	2188	EST01260	M79113	HHCPJ146	2254	EST01315	M79171	HHCPN216	2309	EST01319	M79172	HHCPN50
2123	EST01204	M79053	HHCPH84	2189	EST01261	M79114	HHCPJ148	2255	EST01316	M79172	HHCPN217	2310	EST01320	M79173	HHCPN50
2124	EST01205	M79054	HHCPH85	2190	EST01262	M79115	HHCPJ150	2256	EST01317	M79173	HHCPN218	2311	EST01321	M79174	HHCPN50
2125	EST01206	M79055	HHCPH86	2191	EST01263	M79116	HHCPJ152	2257	EST01318	M79174	HHCPN219	2312	EST01322	M79175	HHCPN50
2126	EST01207	M79056	HHCPH87	2192	EST01264	M79117	HHCPJ154	2258	EST01319	M79175	HHCPN220	2313	EST01323	M79176	HHCPN50
2127	EST01208	M79057	HHCPH88	2193	EST01265	M79118	HHCPJ156	2259	EST01320	M79176	HHCPN221	2314	EST01324	M79177	HHCPN50
2128	EST01209	M79058	HHCPH89	2194	EST01266	M79119	HHCPJ158	2260	EST01321	M79177	HHCPN222	2315	EST01325	M79178	HHCPN50
2129	EST01210	M79059	HHCPH90	2195	EST01267	M79120	HHCPJ160	2261	EST01322	M79178	HHCPN223	2316	EST01326	M79179	HHCPN50
2130	EST01211	M79060	HHCPH91	2196	EST01268	M79121	HHCPJ162	2262	EST01323	M79179	HHCPN224	2317	EST01327	M79180	HHCPN50
2131	EST01212	M79061	HHCPH92	2197	EST01269	M79122	HHCPJ164	2263	EST01324	M79180	HHCPN225	2318	EST01328	M79181	HHCPN50
2132	EST01213	M79062	HHCPH93	2198	EST01270	M79123	HHCPJ166	2264	EST01325	M79181	HHCPN226	2319	EST01329	M79182	HHCPN50
2133	EST01214	M79063	HHCPH94	2199	EST01271	M79124	HHCPJ168	2265	EST01326	M79182	HHCPN227	2320	EST01330	M79183	HHCPN50
2134	EST01215	M79064	HHCPH95	2200	EST01272	M79125	HHCPJ170	2266	EST01327	M79183	HHCPN228	2321	EST01331	M79184	HHCPN50
2135	EST01216	M79065	HHCPH96	2201	EST01273	M79126	HHCPJ172	2267	EST01328	M79184	HHCPN229	2322	EST01332	M79185	HHCPN50

SUBSTITUTE SHEET

-109-

SEQ ID	EST#	Gene	Clone
2393	EST01116	M79261	HRBAA06
2394	EST01117	M79262	HRBAA07
2395	EST01118	M79263	HRBAA20
2396	EST01119	M79264	HRBAA26
2397	EST01120	M79265	HRBAA27
2398	EST01121	M79266	HRBBA04
2399	EST01122	M79267	HRBBA06
2400	EST01123	M79268	HRBBA08
2401	EST01124	M79269	HRBBA11
2402	EST01125	M79270	HRBBA21
2403	EST01126	M79271	HRBBA22
2404	EST01127	M79272	HRBBA23
2405	EST01128	M86176	HRBAA27
2406	EST02114	M86177	HRBAA23
2407	EST00273	M86182	HRBAA23
2408			HRBAA23
2409			HRBAA23

SEQ ID	EST#	Gene	Clone
2398	EST01121	M79198	HRBAA06
2399	EST01122	M79206	HRBAA07
2400	EST01123	M79207	HRBAA20
2401	EST01124	M79208	HRBAA26
2402	EST01125	M79209	HRBAA27
2403	EST01126	M79210	HRBBA04
2404	EST01127	M79211	HRBBA06
2405	EST01128	M79212	HRBBA08
2406	EST01129	M79213	HRBBA11
2407	EST01130	M79214	HRBBA21
2408	EST01131	M79215	HRBBA22
2409	EST01132	M79216	HRBBA23
2410	EST01133	M79217	HRBBA27
2411	EST01134	M79218	HRBBA23
2412	EST01135	M79219	HRBBA23
2413	EST01136	M79220	HRBBA23
2414	EST01137	M79221	HRBBA23
2415	EST01138	M79222	HRBBA23
2416	EST01139	M79223	HRBBA23
2417	EST01140	M79224	HRBBA23
2418	EST01141	M86176	HRBBA23
2419	EST01142	M79225	HRBBA23
2420	EST01143	M79226	HRBBA23
2421	EST01144	M79227	HRBBA23
2422	EST01145	M79228	HRBBA23
2423	EST01146	M79229	HRBBA23
2424	EST01147	M79230	HRBBA23
2425	EST01148	M79231	HRBBA23
2426	EST01149	M79232	HRBBA23
2427	EST01150	M79233	HRBBA23
2428	EST01151	M79234	HRBBA23
2429	EST01152	M79235	HRBBA23
2430	EST01153	M79236	HRBBA23
2431	EST01154	M79237	HRBBA23
2432	EST01155	M79238	HRBBA23
2433	EST01156	M79239	HRBBA23
2434	EST01157	M79240	HRBBA23
2435	EST01158	M79241	HRBBA23
2436	EST01159	M79242	HRBBA23
2437	EST01160	M79243	HRBBA23
2438	EST01161	M79244	HRBBA23
2439	EST01162	M79245	HRBBA23
2440	EST01163	M79246	HRBBA23
2441	EST01164	M79247	HRBBA23
2442	EST01165	M79248	HRBBA23
2443	EST01166	M79249	HRBBA23
2444	EST01167	M79250	HRBBA23
2445	EST01168	M79251	HRBBA23
2446	EST01169	M79252	HRBBA23
2447	EST01170	M79253	HRBBA23
2448	EST01171	M79254	HRBBA23
2449	EST01172	M79255	HRBBA23
2450	EST01173	M79256	HRBBA23
2451	EST01174	M79257	HRBBA23
2452	EST01175	M79258	HRBBA23
2453	EST01176	M79259	HRBBA23
2454	EST01177	M86177	HRBBA23
2455	EST01178	M86178	HRBBA23

SEQ ID	EST#	Gene	Clone
2259	EST01304	M79156	HRBAA22
2260	EST01305	M79157	HRBAA22
2261	EST01306	M79158	HRBAA22
2262	EST01307	M79159	HRBAA22
2263	EST01308	M79160	HRBAA22
2264	EST01309	M79161	HRBAA22
2265	EST01310	M79162	HRBAA22
2266	EST01311	M79163	HRBAA22
2267	EST01312	M79164	HRBAA22
2268	EST01313	M79165	HRBAA22
2269	EST01314	M79166	HRBAA22
2270	EST01315	M79167	HRBAA22
2271	EST01316	M86172	HRBAA22
2272	EST01317	M79168	HRBAA22
2273	EST01318	M79169	HRBAA22
2274	EST01319	M79170	HRBAA22
2275	EST01320	M79171	HRBAA22
2276	EST01321	M79172	HRBAA22
2277	EST01322	M79173	HRBAA22
2278	EST01323	M79174	HRBAA22
2279	EST01324	M79175	HRBAA22
2280	EST01325	M79176	HRBAA22
2281	EST01326	M79177	HRBAA22
2282	EST01327	M79178	HRBAA22
2283	EST01328	M79179	HRBAA22
2284	EST01329	M79180	HRBAA22
2285	EST01330	M79181	HRBAA22
2286	EST01331	M79182	HRBAA22
2287	EST01332	M79183	HRBAA22
2288	EST01333	M79184	HRBAA22
2289	EST01334	M79185	HRBAA22
2290	EST01335	M79186	HRBAA22
2291	EST01336	M79187	HRBAA22
2292	EST01337	M79188	HRBAA22
2293	EST01338	M79189	HRBAA22
2294	EST01339	M86173	HRBAA22
2295	EST01340	M79191	HRBAA22
2296	EST01341	M79192	HRBAA22
2297	EST01342	M79193	HRBAA22
2298	EST01343	M79194	HRBAA22
2299	EST01344	M79195	HRBAA22
2300	EST01345	M79196	HRBAA22
2301	EST01346	M79197	HRBAA22
2302	EST01347	M79198	HRBAA22
2303	EST01348	M79199	HRBAA22
2304	EST01349	M79200	HRBAA22
2305	EST01350	M79201	HRBAA22
2306	EST01351	M79202	HRBAA22
2307	EST01352	M79203	HRBAA22
2308	EST01353	M79204	HRBAA22
2309	EST01354	M79205	HRBAA22
2310	EST01355	M79206	HRBAA22
2311	EST01356	M79207	HRBAA22
2312	EST01357	M79208	HRBAA22
2313	EST01358	M79209	HRBAA22
2314	EST01359	M79210	HRBAA22
2315	EST01360	M79211	HRBAA22
2316	EST01361	M79212	HRBAA22
2317	EST01362	M79213	HRBAA22
2318	EST01363	M79214	HRBAA22
2319	EST01364	M79215	HRBAA22
2320	EST01365	M79216	HRBAA22
2321	EST01366	M79217	HRBAA22
2322	EST01367	M79218	HRBAA22
2323	EST01368	M79219	HRBAA22
2324	EST01369	M79220	HRBAA22
2325	EST01370	M79221	HRBAA22

SUBSTITUTE SHEET

-110-

NOTE REGARDING SEQUENCE LISTINGS: The listings of SEQ ID NOS: 1-2421 are in numerical order. However, an occasional number (for example, SEQ ID NO: 44) is not found in this list. In all, 9 SEQ ID NOS are not used. Nevertheless, the convention "1-2421" is used, for example, to refer to all the SEQ ID NOS in the following list, while "1-315" is used, for example, to refer to all the listed sequences falling between SEQ ID NO 1 and SEQ ID NO 315.

-111-

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Venter, J. Craig
Adams, Mark D.
Moreno, Ruben F.
- (ii) TITLE OF INVENTION: Sequences Characteristic of Human Gene
Transcription Product
- (iii) NUMBER OF SEQUENCES: 2412 (1-2421, with 9 SEQ ID NOS unused.)
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Knobbe, Martens, Olson, and Bear
 - (B) STREET: 620 Newport Center Dr. Sixteenth Floor
 - (C) CITY: Newport Beach
 - (D) STATE: CA
 - (E) COUNTRY: USA
 - (F) ZIP: 92660
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: 07/837,195
 - (B) FILING DATE: 12-FEB-1992
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/716,831
 - (B) FILING DATE: 20-JUN-1991
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Israelsen, Ned A.
 - (B) REGISTRATION NUMBER: 29,655
 - (C) REFERENCE/DOCKET NUMBER: NIH004.004CP1
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 619-235-8550
 - (B) TELEFAX: 619-235-0176

SEQ ID NO:1: (Length of Sequence = 362 Nucleotides)

CTTCCCTTT GTTCCCTCA GTTCCTTT TAATGCTTC CCTCATTTC CCTAGCAGC ATCCTAGTTG ATGGTCTGGG
TTATCAGAGG AGCAAAAACA TTAAAGTGTG AAATAATGCT CATGTCTCC CTGGGATTTC TAAACAGAAA AAATGAAGAA

418

CACTACGGCG GGTGGTCAG CTCATCCCC CTCATCTAG ACTTAATGAA AGAATGGWTC GCCACTTCG AGAAGTTGCC
GCGAAGGTG CTGCAGGGCA CGGGGCTGC CTGCACT

SEQ ID NO:1931: (Length of Sequence = 343 Nucleotides)

ATCACTTCCC CACCCACAG GATCTGCCCC AGAGAAAGTC CTGCTGTC ACCAGCAAGC TTGGGGTGG CCAAGTTGAA
TGATGCTGCC CGGGCTCTG CCAGATCTG AGACGCTTC CCTCCCTGOC CCACCCGGGT CCTGTGCTGG NTCCTGCCCC
TTCTGCTTT TGACGCCAGG GGTGAGGAG TGGCTCGGT GTGGCTGGA GAGGCAGAAG CCTTTCCTG TTGGTGTCCC
AGCATATGA GCCCTTGGG CTGAGCACA AGACCTTGAA CCTTTTTGT TTTACCTTTT TTCCAATAA CAGTTTGGAG
AAATATCAAT GAAATCTGGG GGT

SEQ ID NO:1932: (Length of Sequence = 314 Nucleotides)

TTTCATGGT TTTGTTTT TTTATTCGA ATACTGAAA AGTCCTTGG GCTCTGTGG GTTCCCCAG CTCAGGCTC
CTTCTCCA CACTACTGC CCTTCTCCC ACAGCAAATC TATTCAAGG ACAGTACTTT TTAATATGAT TAATGTGAG
TTCTCACTA GCTCTGAGA ACTAGAGGAG CTGTTTCAT CTGCTGTGC GGATGGAGT TCTTTTATCT GACACCGGT
CTCCAACCAC ACTGATGCAA GGCATTTAT CTACAGAGCT CACTAGAAC CCTTTTCA TTAGGCTACT CCAA

SEQ ID NO:1933: (Length of Sequence = 378 Nucleotides)

AGCTTCTGC GGGACCACAG CTATGTGACT GAAGCTGACA TCATCTCTAC CGTTGAGTTC AACCACAGG GAGAGCTGCT
GGCCACAGG GACAAGGGCG GCGGGTGT CATCTCCAG CGGAACAG AGAGTAAAA TGCGCCAC AGCCAGGGCG
AATACGAGT GTACAGCACT TTCCAGAGCC ACGAGCGGA GTTGTACTAT CTCAAGAGCC TGGAGATAGA GGAGAAGATC
AACAAGATCA AGTGGCTCCC ACAGCAGAAC GCGCCCACT CACTCTGTT CCACCAAGA TAAACTATC AAATTATGA
AGATTACGA ACGAGATAA AGGCCGAG GATACAACT GAAGGATGA GAGGGGA

SEQ ID NO:1934: (Length of Sequence = 239 Nucleotides)

ATTAAATTG ACAGCCTTC ATTTTTOGAG AAAGTACAA CAGAACTGCT TTAGCACCA TOGAGCCCA AACGGGTAG
GTAAGCCAAG GTTTTAATGA CCAGCCAGT ATCTAAGCTT CCAAACGAT GCCAGCCAT CACATACTYA CCTGGGAGG
CTGCTGACG GGCATTCTC YGATGCTAC GGCATTGK GTAGTTTCA RGATGCTC TTTGAGGAAG GACTTCAGG

SEQ ID NO:1935: (Length of Sequence = 319 Nucleotides)

TTAATTTTT TINTCCATAG AGGAATAGCA TTACAGTCTA ACAATCAGAA TTCTGTTACA CACATACACA GGCATGCCAC
ATGACCCAGT TGAGGTGGT GINTCCTGA GTCTGTGAC ACGTCACATG GTCAAAGTCT CCTCATTTCA GCCAGTCTCA
ACACAAAACA CCAACAGGG ATGCACTCAA CTGTGTGGT CCATGTGGA CTAGGTGGCA GGGCAGAGG GAAAGTAGTA
GAAGGGGCT ATGGTGTGTC TGCACTCAGT CCCCTCATAT AAAGCCAT GGATCTAGG GGGGTATCCA AGAGCTCTG

SEQ ID NO:1936: (Length of Sequence = 415 Nucleotides)

CTATTTTAC AAATTATACC TAATGAGTAA AATTAGTGA AAGTGATAAC ATGCTTCTAC CTGTATTCT AGTGACCTT
TAGGGCAGG TATTTATACC TGGTATTAT GATGCAGTAT ATAAGTGGT AACATAACT GACAGTATG TGCTTGCTGT
ACATGCTGG TCTTTGAAA CAGATTTAG TAAGCATTT CCAGAGGTAA AACTGTGTC TTATTCTAAT TTTATCTCA
GGGCAAAGTA GACAGGGATT ATTCCTTGA ATCTATTCC AAATTAATAT TTTTCTTT GGTATTCTA CACTTTAAGG
CCATTGGTG CAATTTAGAA AGTGTGGC TCCTTCCG TAGCCACAT CAAATTAAC TTCCAAACC TCAGGAACAG
TACAAGGAAT TTGAA